Temporal Binding of Action and its Consequences

Stimulus-identity prediction and neural correlates

Tommy Sinnes

Main thesis – Department of Psychology

UNIVERSITY of OSLO

April 2014

Temporal Binding of Action and its Consequences – *Stimulus-identity prediction and neural correlates.*

Year: 2014

Title: Temporal binding of action and its consequences – Stimulus-identity prediction and neural correlates.

Author: Tommy Sinnes

http://www.duo.uio.no/

Abstract

Author: Tommy Sinnes
Supervisor: Tor Endestad
Title: Temporal binding of action and its consequences – Stimulus-identity prediction and neural correlates.

For over a decade, it has been known that intended actions and their sensory consequences are judged as occurring closer together in time than otherwise equivalent non-intended actions and their effects. This phenomenon, known as temporal binding, has been argued to represent a sense of agency and to be the result of motor-predictive mechanisms underlying voluntary action. It has further been demonstrated that such binding can be affected by learning. Based on these findings, I here build on the motor-predictive framework and complementary theories of learning and argue for a central role for prediction error in temporal binding. A computational framework for estimating on-line task-irrelevant predictions in a binding task is proposed and tested on 17 participants while they are simultaneously undergoing functional magnetic resonance imaging (fMRI). It was hypothesised that predictions would be positively associated with degree of binding while the actual outcome would interact with the prediction and form a negative association in terms of increased prediction error. I further speculated that the degree of prediction error would be represented by neural activity in areas more commonly associated with reinforcement learning. The behavioural analysis showed that prediction error was significantly predictive of decreased binding, as hypothesised, but no significant independent contribution of the prediction regardless of outcome was found. For the fMRI analysis, no significant activation was found to be related to prediction error or the prediction itself. However, a second analysis looking for correlations with binding measures revealed negatively correlated activation in the precuneus and left upper brain-stem. The results are discussed in light of the binding measure currently used and a potential role for the default-mode network in agency. It is concluded that task-irrelevant learning of stimulus-identity seems to occur in temporal binding, and seems driven by prediction error. It is, however, still unknown how this is implemented in neural terms. Because the current study fails to replicate an earlier finding regarding neural correlates of temporal binding, a continued investigation of this is encouraged, as well as further studies on how different binding measures relate to agency and voluntary action.

Preface

The current study has been devised and conducted by the author. This includes theoretical preparation, experimental design and related programming, recruiting participants, data collection and subsequent statistical analysis. Associate professor Tor Endestad has supervised the process and provided comments on the design, analysis, and the manuscript. Data collection was performed by the author and a scanner assistant between November 2013 and March 2014 at the Intervention Centre at Oslo University Hospital, Rikshospitalet.

I would like to thank all participants for graciously lending me their time, and being very understanding even at times when there were delays occurring. Their good effort is highly appreciated. I would also like to thank Lars Christian Vold for assisting during data collection, as well as Grethe Løvland and Svein Are Vatnehol for being available for radiographic assistance when needed. Last, but not least, I would like to thank Tor Endestad for giving me this great opportunity to conduct the study and being highly available for supervision whenever needed.

Table of contents

1 Introduction

1.1 Agency, mechanisms of motor control, and the experience of time

"*It is still open for me, as well as you, to regulate my behavior, by my experience of past events."* (Hume, 1758, p. 361)

As human agents we enjoy the ability to influence our environment through our actions, allowing us to achieve desired future states in relation to that environment. Such an ability requires modelling of the environment (Conant & Ross Ashby, 1970), which in turn requires information to be gathered regarding the models success or failure in order for it to adapt according to environmental changes. If a change occurs, it must be known whether the change could be attributed to the agent. In other words, some inference about causality between the agent and the environmental change needs to be made. This involves the inference that 'the action caused the effect', and, perhaps more fundamentally, the inference that 'I caused the action'. The ability to refer to oneself as the author one's own actions, has been regarded the defining feature of the human sense of agency (de Vignemont & Fourneret, 2004), which further involves an experience of coherence between the intended action and its perceived consequence. As has been shown quite recently, this experienced coherence seems not to be merely an abstract experience, but seems to manifest itself in very concrete perceptual terms where the action and its consequence are actually perceived as being bound together in time (Haggard, Clark, & Kalogeras, 2002), a finding that will shortly be discussed in more detail. Intuitively, it may thus seem as (experienced) agency, motor control and the perception of time are closely related phenomena.

Before proceeding with a more detailed discussion of the binding phenomenon, it should be noted that there is a fundamental problem in measuring subjective time in that the experienced time has to be reported after the occurrence of the event, necessarily involving both perceptual and recollective processes. Findings of reliable differences in such reports, e.g. between conditions in an experiment, could thus be due to recollection bias as well as an actual change in perception. The binding phenomenon, that involves a reported shortening of the interval between action and effect, has nevertheless been thought to be the result of a

direct influence on perception (e.g. Haggard et al., 2002), similarly to the process of inferring causality in visual perception (Körding et al., 2007; Scholl & Nakayama, 2002; Scholl & Tremoulet, 2000). This assumption follows largely from findings linking binding specifically to voluntary action (Haggard, 2005), which in turn implies a central involvement of mechanisms underlying motor control. Some understanding of these mechanisms is thus required.

1.1.1 Motor control and sensory processing

One of the most influential computational frameworks of the neuroscience of voluntary action (Wolpert, 1997; Wolpert & Ghahramani, 2000) holds that this is governed by 'internal models' modelling the sensorimotor system, of which there are three types: 'inverse models', 'forward dynamic models', and 'forward sensory models'. Specifically, this framework states that, given the state of the sensorimotor system in combination with a task to be executed, a motor command is specified (inverse model). The expected change in the sensorimotor state is then computed from the motor command (forward dynamic model) and specific sensory predictions are generated given this expected state (forward sensory model) (fig. 1). The difference between predicted and actual sensory feedback provides an error signal that can be used to update the model. In this way, new associations between desired sensory outcomes and the actions required to generate them, can be learned. The internal sensory predictions are further considered a central component of our normal experience of control or agency. Here, predictions are thought to be compared with actual sensory consequences and a sense of agency arises as a result of how well they are matching. This explanatory framework has been termed the 'comparator model' and has been used to explain normal action awareness (Blakemore, 2009), as well as delusions of control (Blakemore, Wolpert, & Frith, 2002; Frith, 2012). Such a sense of agency is important because it allows differentiating between sensory changes that are caused by the agent and changes that are externally generated. The internal sensory predictions are thought to affect the perceptual system directly, resulting in phenomena like 'sensory attenuation', where sensations are experienced as less intense after being predicted by motor processes (e.g. Bays, Flanagan, & Wolpert, 2006; Blakemore, Wolpert, & Frith, 2000; Cardoso-Leite, Mamassian, Schütz-Bosbach, & Waszak, 2010; Voss, Ingram, Haggard, & Wolpert, 2006). Direct evidence for such modulation of the sensory system has been provided (Hughes & Waszak, 2011; Shergill et al., 2013) and has been explained in terms of 'preactivation' of sensory activity (Roussel, Hughes, & Waszak, 2013). Sensory prediction has also been linked to explicit agency

judgments in experimental studies (Sato, 2009; Sato & Yasuda, 2005), supporting its role in generating a sense of agency.

Because the phenomenon of temporal binding has been observed in close relation to voluntary actions, it has been assumed that internal sensory predictions are correlated with the phenomenon, and it has therefore been suggested that they are actually causing it. Since these sensory predictions are occurring before the actual sensation is experienced and involve processes in the perceptual system, it is thought that temporal binding could also be a perceptual phenomenon. The sensory effect component of the action-effect interval has specifically been explained in terms of sensory preactivation (Desantis, Hughes, & Waszak, 2012): The sensation is evoked before its generative external event actually occurs, and is therefore more readily experienced and appears to occur earlier in time than what is actually the case.

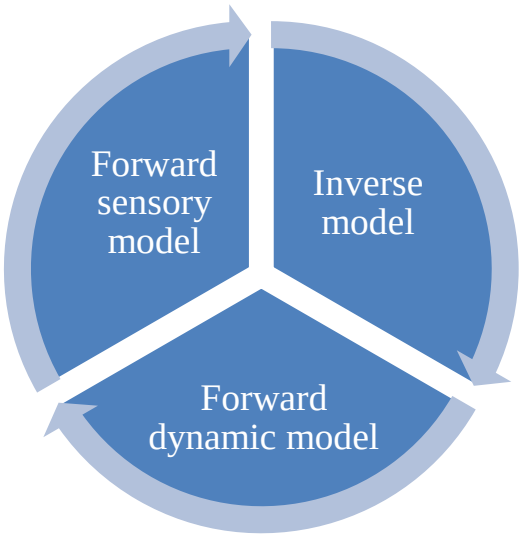


Figure 1. *The stages defining the process of generating movement. The model involves continuous updating through a specific mechanism which is not illustrated here, but occurs through the use of the above described components. The figure is based on Wolpert and Ghahramani (2000).*

The assumed importance of motor prediction in binding does of course not rule out the possibility of an involvement of recollective processes, it merely states the potential for predictions to affect perception during voluntary generated action-effects. With this in mind, it should also be noted that alternative models of experienced motor control holds that the experience of motor control or agency is essentially a reconstruction of likely causation that is

happening after the perceived influence of action (Wegner, 2003). However, the potential involvement of postdictive processes does not necessarily imply high-level cognition either, since it has been shown that such postdictive factors can actually influence low-level perception (e.g. Eagleman & Sejnowski, 2000; Eagleman & Sejnowski, 2003) and that it can also affect causal perception (Choi & Scholl, 2006).

More recently, an alternative account of action and perception has emerged based on hierarchical Bayesian inference (Friston, 2010; Friston, Kilner, & Harrison, 2006). From this, it is argued that both action and perception arises out of the same principle of trying to minimise discrepancy between sensory evidence and predictions (Adams, Shipp, & Friston, 2013; Friston, Daunizeau, Kilner, & Kiebel, 2010), and further, that the same principles underlies intentions and understanding of other peoples actions (Friston, Mattout, & Kilner, 2011; Kilner, Friston, & Frith, 2007). I will be returning to this subject of minimising discrepancy, or prediction error, later.

Taken together, theories and empirical findings related to mechanisms of motor control suggest a central role for sensory prediction in action and perception. These predictions are further thought to be fundamentally involved in generating a sense of agency. The idea that time perception could be equally affected by these same principles underlying motor control, paves the way for the intriguing possibility that the experience of coherence, or tieing-together, associated with agency experience could have a clear perceptual basis. It follows easily from this, that further understanding of the common mechanisms underlying these phenomena could have the potential to shed new light on the phenomenology of agency. Furthermore, this is not merely of philosophical interest, but is thought to be of special importance in understanding abnormalities in the experience of agency. For patients diagnosed with schizophrenia there are not necessarily obvious problems with basic motor control, but still they frequently report subjective experiences of vastly excessive control or a complete lack of control when this is evidently not the case. Keeping this motor control framework in mind, I will now turn to more specific findings on the variability of experienced time of action-effect intervals.

1.2 Temporal binding of action and sensory consequence

1.2.1 The intentional binding effect – intention, causality or agency?

The finding that voluntary actions and their sensory consequences are perceived as temporally shifted together in time was first demonstrated in a study by Haggard et al. (2002).

By watching a rotating clock and pushing a button at a time of their choosing, participants reported either the time of their button press or a resulting tone, following 250 ms after the action. The reported times were compared with judgments of equivalent actions and tones occurring alone. In a separate condition, involuntary muscle twitches were induced by applying transcranial magnetic stimulation (TMS) to the motor cortex. These twitches were also followed by tones, and both components were judged in the same manner as in the voluntary condition and compared with twitches and tones occurring alone. Additionally, a separate sham TMS condition was included involving similar comparisons. For both the voluntary and involuntary action condition, the judgments were significantly different when action and tone occurred in sequence. However, they showed opposite patterns of temporal shifts. Whereas the action in the voluntary condition was perceived as occurring later in time, the action in the involuntary condition was perceived as occurring earlier. The opposite pattern was demonstrated for tone judgments in the voluntary and involuntary condition. Based on these findings, it was argued that the temporal binding of actions and effects is specific to intentionally generated actions and was therefore termed 'the intentional binding (IB) effect'. It was further speculated that the phenomenon represents a construction of a coherent conscious experience of agency by binding intentional actions to their effects. In a subsequent study, the effect was also demonstrated when participants were observing other peoples intended actions, compared with machine-generated actions (Wohlschläger, Haggard, Gesierich, & Prinz, 2003), supporting an involvement of intention.

The assumption that intention, and not a more general condition of causality, was the crucial ingredient underlying IB was however soon challenged (Eagleman & Holcombe, 2002). The importance of perceived causality for IB to occur has been supported by more recent empirical studies (Buehner, 2012; Buehner & Humphreys, 2009). However, Moore and Obhi (2012) and Cravo, Claessens, and Baldo (2009) have argued for both causality and voluntary action being necessary for the effect to occur. Nonetheless, the phenomenon of action-effect binding will in the following be referred to using the more neutral term temporal binding (TB), not implying a necessary involvement of intention.

Regardless of true intentionality being a necessary component of TB, it has been assumed that the binding phenomenon has a special relation to the experience of agency (Moore & Obhi, 2012). In fact, this assumption has been so strong that several studies have used TB as an indirect measure of sense of agency (e.g. Demanet, Muhle-Karbe, Lynn, Blotenberg, & Brass, 2013; Engbert, Wohlschläger, & Haggard, 2008; Kühn, Brass, & Haggard, 2013). Importantly, this experience could represent true agency resulting from

motor-predictive processes, but it could equally represent a false inference of agency based on postdictive factors.

1.2.2 Predictive mechanisms in TB

Because of its apparent link to voluntary, intended actions, the TB effect was early hypothesised to be a consequence of sensory predictions resulting from forward sensory models during action (Haggard et al., 2002). Consistent with this hypothesis, there has been found evidence of altered binding in patients with schizophrenia (Haggard, Martin, Taylor-Clarke, Jeannerod, & Franck, 2003; Voss et al., 2010), a patient group in which the disrupted sense of agency has previously been argued as resulting from abnormalities in internal forward models (Blakemore et al., 2002). This hypothesis of abnormalities in action awareness in schizophrenia has received additional evidence in recent years (Lindner, Thier, Kircher, Haarmeier, & Leube, 2005; Synofzik, Thier, Leube, Schlotterbeck, & Lindner, 2010; Williams, Ramachandran, Hubbard, Braff, & Light, 2010).

Looking for experimental evidence of such predictive mechanisms in binding, Tsakiris and Haggard (2003) replicated the intentional binding effect for a somatosensory consequence, and showed in a separate experiment that the same consequence also was perceived as less intense when occurring after a voluntary action, in line with the sensory attenuation effect previously described. This is indicative of sensory predictive mechanisms operating during a binding task, but does not imply a causal relationship between the two. As a continuation of the initial investigations into TB, Haggard and Clark (2003) further demonstrated that TB is decreased when an intention is disrupted by an involuntary movement, even though it is followed by the same sensory effect, implying that retrospective inference is not sufficient as an explanation of the phenomenon, and that internally generated predictions are crucial.

In further support of predictive mechanisms, it has been shown that priming of effect increases binding when the prime is congruent with the effect (Moore, Wegner, & Haggard, 2009). This effect of priming was found for both voluntary and involuntary action-effect relations, but was actually strongest for involuntary actions. Such an influence is also consistent with the view that TB relates to experience of agency, as this experience has also been found to be influenced by priming (Aarts, Custers, & Wegner, 2005).

The sensory prediction view on TB was, however challenged quite recently, in a paper by Desantis et al. (2012). There, it was pointed out that previous studies had not made sufficient distinctions between all relevant variables and that conclusions regarding TB effects

due to specific sensory predictive processes were not warranted. An experiment was described where the effect of identity-specific predictions on TB was separated from factors like motoric control of the onset of the stimulus. A classic TB effect was demonstrated, but no specific effect was found for the ability to predict exact stimulus-identity. It was then concluded that specific sensory prediction plays no role in TB. Because of importance for the assumptions from which the current study is designed, this claim should be commented further.

There are some critical comments that could be made regarding this finding. First, the study assessed binding using an indirect approach (see later paragraph 'Measuring temporal binding' for further explanation) and in so doing, only considered the temporal shift of the action-effect. This was chosen because it was thought to be more dependent on predictive processes than the action, based on a previous report that showed effect of pre-SMA stimulation on effect, but not action judgments (Moore, Ruge, Wenke, Rothwell, & Haggard, 2010). However, we do not know whether a direct interval measure would have produced different results. Second, the study reports on a failure to reject the null hypothesis under classical inference and then draws conclusions about the truth of that hypothesis, which is strictly wrong (Gallistel, 2009; Wagenmakers, 2007). Still, the finding that TB seems to be driven by other factors like temporal control under those experimental circumstances should be considered important. With this in mind, perhaps the most central critique pertains to the fact that the study is built on the premise that sensory predictions resulting from forward models provide either a match or a no-match with respect to the output in a qualitative-like fashion. Specifically, it is claimed that "if predictive forward mechanisms drive binding, then it should only occur in situations in which the agent is able to predict the identity of the sensory event s/he is going to produce" (Desantis et al., 2012, p.1). In contrast, the discrepancy between predicted and actual sensory consequences are considered to be quantitative in nature (Blakemore, 2009) stated in terms of prediction error (Wolpert & Ghahramani, 2000). This matters because the sensory consequences used in the study of Desantis et al. (2012) are two tones, differing only in pitch. Under a forward model where neural representations of the predicted stimulus are compared against the actual stimulus representation, the instance where the predicted stimulus are completely absent should only be regarded as the most extreme violation of the prediction for that particular stimulus. An instance where the predicted stimulus is somewhat wrong could still be viewed as a quite close match. Thus, to have the best chance of getting an influence on TB from violations of predicted stimulus identity, the alternative identities should be as different as possible in terms

of evoked neural representations. The findings of Desantis et al. (2012) are therefore not considered convincing for rejecting a contribution of stimulus prediction in TB.

Not only internal predictive mechanisms have been found to exert an effect on binding. It has been demonstrated that what is happening after the prediction, so-called postdictive factors, also play a significant role in influencing the magnitude of TB. An 'optimal cue integration model' has been proposed by Moore et al. (2009) to explain the combined influence of motor signals and sensations evoked by external cues on TB (see also Moore & Fletcher, 2012). Some of the findings relating to the role of postdiction will now be reviewed from a learning perspective.

1.2.3 Learning effects on TB

An implication of forward models, or any model involving sensory prediction, is the potential for such predictions to be learnt. If such predictive mechanisms are involved in TB, one could expect that those mechanisms would indeed be under the influence of some learning process.

In line with this, Engbert and Wohlschläger (2007) found that binding of the action component was dependent on the probabilistic contingency between action and outcome irrespective of whether the outcome actually occurred, indicating that this probabilistic association was learnt and in turn affected TB. Similarly, Moore and Haggard (2008) demonstrated that when no outcome appeared, action binding was increased in a condition where expectancy regarding outcome was higher. Further, they showed that in a condition where occurrence of the stimulus was unpredictable, binding was increased when the stimulus did occur. Thus, it was argued that both predictive and retrospective mechanisms contribute to TB. Importantly, these findings imply that TB is not fixed due to some solely internal process, but are affected by external cues through both immediate corrections of likely causation and updating of future expectations. However, as pointed out by Desantis et al. (2012), the studies are concerned with prediction of stimulus occurrence, and not with prediction of the exact identity of the stimulus.

Of particular importance for the current study is the additional finding in the study of Moore and Haggard (2008) that the effect on binding in the prediction condition was modulated by recency of outcome. That is, when an action had recently produced an outcome, binding would be increased on subsequent trials. This indicates that predictions are continuously updated according to learning principles similar to those of basic associative learning. Such learning principles were investigated further by Moore, Dickinson, and

Fletcher (2011). Specifically, they measured binding for action-effect relations that had been previously trained under different levels of surprise and found increased binding when the surprise level had been higher during training. This indicates that surprise plays a role in establishing future expectations which in turn affects degree of binding.

In a study by Aarts et al. (2012) it was found that priming people with positive, compared to neutral, pictures increased binding in a subsequent TB task. This was thought to be related to reward processing. Perhaps even more interesting, this effect was found to be mediated by the participants' baseline eye-blink rate as measured by eye-tracking equipment. This was considered indicative of a potential role for dopamine in TB because of its presumed association with eye-blink rate (Karson, 1983; but see also van der Post, de Waal, de Kam, Cohen, & van Gerven, 2004). More direct evidence for dopaminergic influence on TB was however provided by Moore, Schneider, et al. (2010) who demonstrated increased binding in patients with Parkinson's disease when on dopaminergic medications compared with binding when off such medications. Since dopamine function is also altered in schizophrenia (e.g. Howes & Kapur, 2009; Toda & Abi-Dargham, 2007), one might further speculate on whether the previously mentioned findings of increased TB in these patients also relates to alterations in dopamine. Taken together, these findings provide an even stronger case for associative learning processes in TB given the central role of dopamine in basic reward learning (e.g. Wise & Rompre, 1989) and possibly also in the more general context of associative learning (Spanagel & Weiss, 1999).

Even though a great amount of studies imply a role for learning in TB, there is a lack of studies investigating the specific mechanisms underlying such learning. Seeking to explore such learning mechanisms further, I here draw on previous success in modelling task-irrelevant associative learning using a simple computational framework in combination with a temporal estimation task.

1.3 A computational framework for on-line associative learning in TB

1.3.1 The concept of prediction error in models of learning

From the TB studies reviewed, it is evident that some sort of learning mechanism is involved. Central to both classic models of associative learning (Mackintosh, 1975; Pearce & Hall, 1980; Rescorla & Wagner, 1972), as well as more recent generalised predictive coding models (Friston, 2010; Friston et al., 2006) and previously mentioned models of motor control

(Wolpert & Ghahramani, 2000), is the concept of prediction error, or surprise. It is thought that learning only occurs when the outcome is surprising with regards to some prediction, that is, when there is an error in the stated prediction. More specifically, the prediction is updated according to the magnitude of prediction error. Given the aforementioned association between TB and dopamine, it is then particularly interesting that a central function of the dopamine system seems to be encoding of prediction errors during reinforcement learning (Bayer & Glimcher, 2005; Glimcher, 2011; Pessiglione, Seymour, Flandin, Dolan, & Frith, 2006; Schultz, Dayan, & Montague, 1997). Similar prediction error signals has also been found in the striatum for non-rewarding and task-irrelevant learning of stimulus-stimulus associations (den Ouden, Friston, Daw, McIntosh, & Stephan, 2009), indicating a similar role for prediction errors in associative learning and adding to other findings of learnt task-irrelevant statistical associations (Turk-Browne, Jungé, & Scholl, 2005; Turk-Browne, Scholl, Chun, & Johnson, 2008; Turk-Browne, Scholl, Johnson, & Chun, 2010). Recent theoretical work based on the so-called 'free-energy principle' casts further doubt on the specificity of reinforcement learning (Friston, Daunizeau, & Kiebel, 2009), and redefines the role of dopamine accordingly (Friston et al., 2012). This is based on a previously mentioned hierarchical predictive coding model of the brain where it is thought that each level in the hierarchy seeks to explain input from lower levels by continuously minimising prediction error (Friston, 2010; Friston et al., 2006). Evidence for hierarchical prediction errors has recently been reported (Iglesias et al., 2013).

Seemingly mirroring the link between abnormal TB and schizophrenia, patients suffering from psychosis have also been shown to have abnormal neural responses associated with reward prediction errors (Murray et al., 2007), and patient's with Parkinson's disease off dopaminergic medications have shown decreased responses to prediction error (Galea, Bestmann, Beigi, Jahanshahi, & Rothwell, 2012).

Given that both TB and prediction error are related to learning, dopamine function and schizophrenia, a central role for prediction error signals in driving binding is assumed. Further, it is suspected that these signals might draw on the same mechanisms involved in classic reinforcement learning, and hence any associated neural activity might be suspected to overlap with neuroanatomical regions implicated in processing of reward.

1.3.2 Measuring temporal binding

TB have primarily been assessed using two different approaches (Moore & Obhi, 2012). Following the methodology for measuring subjective time introduced by Libet (Libet,

Gleason, Wrigth, & Pearl, 1983), an indirect procedure was originally used, having participants watch a rotating clock-hand and report its position either when they pressed the button or when the effect appeared (Haggard et al., 2002). The interval judgments for each condition could then be estimated post hoc. Because of the limitations of this method (e.g. Humphreys & Buehner, 2009), a direct interval estimation method was later introduced to the field (Engbert, Wohlschläger, Thomas, & Haggard, 2007). This methodology simply involved having participants report the duration of the interval using a scale.

Some limitations are associated with both procedures. In order to measure the entire action-effect interval at once, a direct estimation method has to be used. Using the indirect procedure therefore also requires twice the amount of trials in order to reconstruct the combined amount of binding resulting from the separately perceived shifts of action and effect. The direct estimation procedure however involves some sort of verbal report where intervals have to be reported using a measure of duration that can not be perceived directly. In other words, a transformation from the experienced time duration to a more abstract time scale is required. Even though the method has proven to reliably reproduce the TB effect (Engbert et al., 2007), this transformation could be expected to introduce additional noise in the measurements. Further, this method has employed only a few distinct alternating interval durations during the task, making the durations more easily predictable over time. Potentially overcoming these limitations, a direct estimation method based on reproduction was recently introduced by Poonian and Cunnington (2013) were the intervals varied randomly between 500 ms and 1500 ms. Here, subjects reported the perceived interval length by simply reproducing the duration through pressing and holding down a button for the same amount of time as the perceived length. Such a motor-based report could be argued to be a much simpler and more experience-near way of reporting subjective time, as well as providing the opportunity for a higher degree of randomness from trial to trial, thereby reducing temporal control of stimulus onset (Desantis et al., 2012).

1.3.3 On-line prediction error changes in TB?

An earlier study has shown that surprise affects later measurements of binding (Moore et al., 2011), but it is currently unknown whether prediction error has an on-line effect on TB in a linear fashion and, if so, how it actually relates to binding in terms of an increasing or decreasing effect. Even though indications of such prediction error effects have been provided through the modulating effect of recency of previous action-effect associations (Moore & Haggard, 2008), there have, to my knowledge, not been any attempts of relating moment-to-

moment changes in prediction error to changes in TB at those same moments. Further, earlier studies have failed to test for effects of specific stimulus-identity predictions (Desantis et al., 2012), which is a crucial implication of motor-predictive mechanisms (Wolpert & Ghahramani, 2000). I therefore aimed to further investigate the combined influence of prediction and the resulting outcome on TB for predictions that were only concerned with stimulus-identity. It was hypothesised that prediction errors would drive learning in terms of updating specific sensory predictions from the motor system, and that these would in turn affect the magnitude of TB depending on the actual outcome. It was suspected that the magnitude of prediction error would be negatively related to the magnitude of TB, as would be expected from a comparator model of agency experience. Because previous studies had shown an independent effect of expectation on TB (Engbert & Wohlschläger, 2007), it was further hypothesised that the prediction level would additionally affect TB regardless of the actual outcome.

In light of recent findings implicating reward processing neuroanatomical regions in learning statistical associations even in the absence of reward (den Ouden et al., 2009), it was thought that such regions might also be implicated in motor-prediction learning underlying TB.

1.3.4 Combining prediction estimates and TB measures

Attempting to investigate both the hypothesised impact of prediction errors and resulting predictions, as well as underlying neural activity, I set out to devise a TB task that would make possible characterisation of such predictions and errors together with measurements of neural activity. When seeking to investigate the role of stimulus prediction in TB, it is of utmost importance that the learning is irrelevant for performance of the TB task. Thus, it follows that the estimation of prediction changes can not be aided by behavioural measures. It was therefore sought to use a theoretical model to compute such predictions based on experimentally manipulated variables. The model would preferably have considerable empirical support and be simple, requiring few additional assumptions to be made.

With these thoughts in mind, I adopted parts of the paradigm and procedure used by den Ouden et al. (2009) in exploring the role of prediction error in incidental learning. An experiment was developed where participants were to freely choose one of two buttons to press, which then in turn was followed by a sensory effect. The length of the interval was then judged, similarly to other TB tasks. However, the button press was then set up to act as a cue

towards the identity of the effect. Following den Ouden et al. (2009), the probability of the two identities were otherwise identical, implying that the only way to reach above-chance predictions regarding the stimulus-identity was to rely on the action-cue. This was achieved by making sure participants made an equal amount of presses on each button while each button was probabilistically associated with the stimulus-identity in a mirrored fashion to the other button. As has been critically remarked previously in the study of TB (Desantis et al., 2012), the hypothesised underlying predictive mechanisms, as specified by forward or predictive coding models, crucially implies stimulus-identity prediction. I therefore sought to devise a temporal judgment task were the action-effect would only differ in terms of a specific feature. Colour was chosen as the stimuli-distinguishing feature because of its ability to evoke visual neural activity (Beauchamp, Haxby, Jennings, & DeYoe, 1999).

In terms of the specific model serving to estimate participants acquiring of predictions, I also chose the Rescorla-Wagner model (Rescorla & Wagner, 1972), used successfully by den Ouden et al. (2009) to model incidental probabilistic learning. This model is considered a classic model of associative learning and has proven fairly accurate in modelling several simple learning phenomena (Miller, Barnet, & Grahame, 1995). The model was not chosen because it was considered to be the best model for stimulus prediction in TB, but because it was thought to be able to model some of the variation due to prediction error-driven learning. Importantly, if the model proves to capture some of the variation in TB, it should be equally capable of describing the effect of prediction level in combination with outcomes as the effect of prediction level regardless of outcome. Since prediction error is defined as the discrepancy between prediction and outcome, the combination of these two variables will define the degree of prediction error. Because the prediction will be defined with respect to the high-probability outcome separately for both cues, the outcome on each trial will determine whether an increasing prediction value represents increasing or decreasing values of prediction error. I will be referring to these two characterisations of prediction error as increasing surprise and increasing expectancy, respectively.

In order to try to capture underlying neural activity associated with model-derived predictive learning, the task was designed to be compatible with a functional magnetic resonance imaging (fMRI) design, similarly to den Ouden et al. (2009). Using fMRI, I wanted to test the hypothesis that sensory predictions during TB could be associated with activation of structures central to reinforcement learning. A control condition to validate the occurrence of a real TB effect was not included in the experiment in order to maintain a reasonable time

frame for the freely volunteering participants and at the same time get adequate power for the fMRI measurements during the TB task.

In this particular design it was considered important to have an equal number of button presses for each hand. The participants were therefore forced to use a fixed number of button presses during a given block of the task, which necessarily resulted in the last trial(s) in each block being completely forced, involving no freedom of choice. However, because of the blocked nature of the design, these few trials would be distributed across the experiment. It was assumed that the effect of these forced trials could be neglected with respect to the overall effect.

Because it was thought to be the easiest and most reliable way of reporting subjective time duration while introducing trial-to-trial randomness in interval lengths, the reproduction procedure of Poonian and Cunnington (2013) was chosen as the method to measure TB. Slightly diverging from this procedure, it was additionally chosen to correct for variance in actual duration and compute an estimate of the relative binding for each trial. This was partly motivated by theoretical considerations of isolating what was thought to be the relevant component of the judgment error, but also because it previously has been shown that these absolute errors increase with interval length when using a direct estimation procedure (Humphreys & Buehner, 2009).

1.4 Current hypotheses

It was hypothesised that an increased prediction regarding stimulus-identity would be associated with increased TB, as evidenced by a shortening of the reported interval durations relative to the actual durations. It is further suspected that the prediction level will interact with the actual stimulus outcome such that an increased level of prediction error will be associated with a decrease in TB.

Further, it is suspected that such prediction error-induced learning involves activation of neuroanatomical regions which are also known to be involved in classic reinforcement learning. By investigating neural correlates of the same model-derived expectancy/surprise values, it is sought to provide indirect evidence for such involvement. Specifically, it is hypothesised that increased levels of surprise will be associated with activation of striatum and certain regions of the visual cortex (because of the surprising feature being visual).

2 Methods

2.1 Participants

A total of 32 healthy volunteers participated in the current study. None of them reported any ongoing neurological or psychiatric disorder and all had normal or corrected-to-normal vision, including normal colour-vision. Participants were either 1st year psychology students or students in other disciplines. They were all informed about their right to leave the study at any time and gave written consent before entering the scan room. A standard MRI safety procedure was also performed for all participants before scanning.

Because of a programming error resulting in a failure to separate left and right button presses in the analysis for the first 8 participants, these participants had to be excluded from the current study. An additional 6 participants did not complete the interval estimation task, either because of too slow responding or misunderstanding of the task instructions. 1 participant also aborted the task because of concerns with the scanner noise. This left me with complete datasets from a total of 17 individuals (8 male) ranging from 20 to 31 years of age (mean = 22.47; SD = 3.50).

2.2 Design

The current study was designed around 3 factors. Evolving prediction (*prediction*: factor 1) was defined according to button press (*cue*: factor 2) acting as a cue for the identity of the stimulus outcome (*outcome*: factor 3), which is further providing information to update the prediction. The *prediction* could vary continuously between 0 and 1, while *cue* and *outcome* each had two distinct levels. The *cue* could be either colour- or grey-predictive, while the *outcome* could be either colour or grey. Importantly, this allowed me to characterise TB and neural activity associated with *prediction* given the *cue*, as well as the interaction between *outcome* and *prediction* given the *cue*, resulting in characterisation of surprise and expectancy, or increasing and decreasing levels of prediction error.

2.2.1 Interval estimation task

As will also be evident from the task instructions below, participants were presented with a fixation point displayed at the centre of the screen throughout the experiment. This point changed between being a vertical-horizontal cross and a simple horizontal bar. When the participants pressed one of the buttons, a circle followed after a uniform randomly chosen

interval between 500 and 1500 ms and was then presented for 500 ms. Apart from being central to the task, this random duration prevented the participants from estimating the exact onset of the stimulus (temporal control) and added an additional jitter which is of advantage to later fMRI analysis. After participants had tried to reproduce the interval by again pressing down the button, the fixation bar appeared upon releasing of the button. A jittered interval (uniform) between 1000 and 4000 ms was then added before the start of a new trial was marked by the reappearance of the fixation cross, indicating that a new button press could be made (fig. 2).

The task was further divided into blocks, a total of 10, wherein participants had to use a total of 20 left and 20 right button presses. The remaining presses for the left and right button were continuously displayed in the upper left and right corners of the screen, respectively. If a button was pressed after all its available presses had been used, a message was displayed at the screen telling the participant to press the other button. Between each block, there was a short break of 12 seconds where only the fixation bar was displayed. At the start of the next block, the counters reappeared, displaying the number 20 in each corner. The task was further divided into two separate scan sessions, containing 5 blocks each, in order to give participants a short break before continuing. The whole task comprised 400 trials of interval estimations, 200 for each action cue.

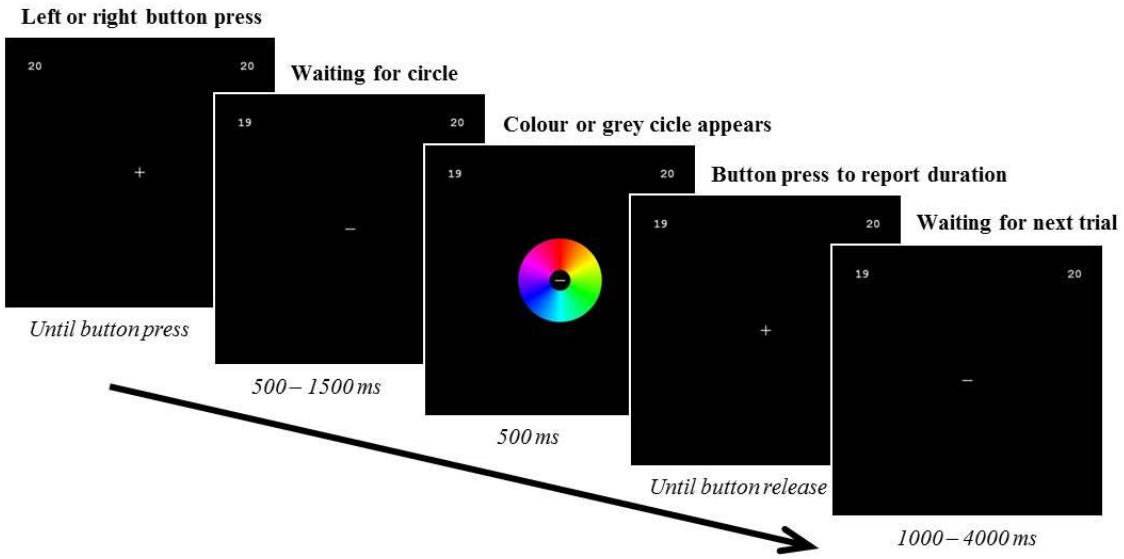


Figure 2. *Illustration of one trial of the task. Fixation cross indicates that a button press should be made and fixation bar serves as a signal to wait. Counters in the upper corners denotes remaining left and right presses for the current block. Bottom text describes the duration of each frame. Proportions are not identical to the actual presentation (fixation points and numbers have been resized and adjusted for illustration purposes).*

2.2.2 Stimuli and task-irrelevant probabilistic associations

As with no relevance for participants in their given task, the circles resulting from a button press were either completely grey or coloured. However, left and right button presses were programmed in such a way that each caused a different amount of grey or coloured circles in each block. Specifically, one button resulted in 80 % coloured stimuli and 20 % grey, while the other button produced the exact opposite pattern (fig. 3). Thus, the stimulus identity was highly contingent on the specific button that was pressed. The proportion of grey to coloured stimuli was fixed for each block, having one button always result in 16 colour and 4 grey circles and the other button getting the exact opposite distribution of stimuli. The order of colour and grey was randomised for each button. Importantly, and as an effect of the two action-outcome contingencies mirroring each other, there was an equal amount of both stimuli in each block, implying button press type as the only available predictor of stimulus identity. To minimise any bias of left and right presses on TB, the buttons contingent on colour and grey was switched across subjects, resulting in 9 participants having the right button contingent on colour and the 8 other having the right contingent on grey.

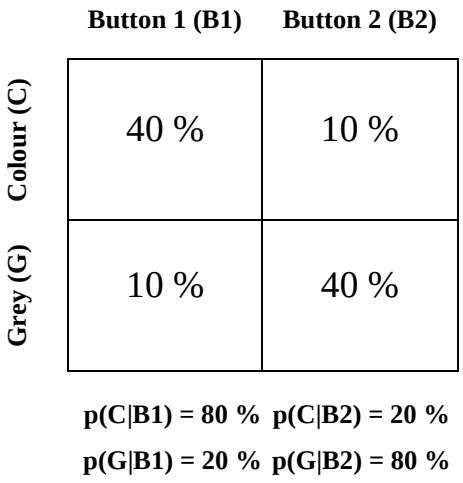


Figure 3. *Contingency table, showing proportion of trials in each block displaying coloured or grey stimuli contingent on the pressing of either of two buttons. Which of button 1 or 2 that was assigned to right or left, differed between participants.*

As colour and grey stimuli I chose colour and luminance-matched greyscale wheels similar to the kind used in fMRI-adapted versions of the Farnsworth-Munsell 100 Hue Test (Beauchamp et al., 1999). Because I was interested in specific stimulus-identity prediction,

and not a more abstract prediction of colour, only one colour wheel was created for each participant. The wheel was created as a standard RGB colour wheel where colours changed gradually from red to green to blue and back to red, in clockwise direction. This colour stimulus was then converted to a grey-scale representation of its luminance via an in-built Matlab function (fig. 4). The stimuli were not adjusted according to subjectively perceived differences in luminance because it was regarded essential that the stimuli were novel at the start of the estimation task, to hinder any influence on the subsequent potential learning process. Further, the aim of the experiment was not to study colour processing per se, so the possibility of some activation being due to perceived luminance differences would not be of importance for the current questions. Following Beauchamp et al. (1999), the size of the wheels were adjusted such that they extended from ~ 1° to ~ 4° of the subjects' field of view according to the central fixation point, because this was presumed to be the part of the retina with the highest density of cones. Across participants, three different variations of the colour and corresponding grey stimuli were used, one starting with red at the 12 o'clock position (fig. 4), and the other two stimuli starting with red at 4 and 8 o'clock, respectively. This resulted in 5, 6, and 6 participants having the respective stimulus variations, and it was done to hinder any bias resulting from the specific orientation of the stimuli.

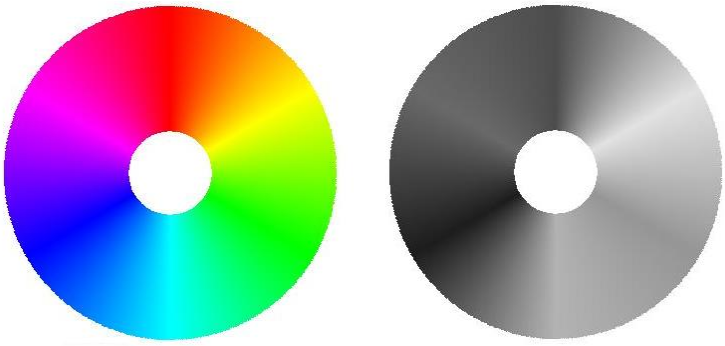


Figure 4. *One set of colour and luminance-matched grey-scale wheels that was used as stimuli in the task. Three differently rotated sets of stimuli were created, and one set was assigned to each participant.*

2.3 Procedure

2.3.1 Task instructions

When volunteering, all participants were told that the aim of the study was to investigate judgments of time and neural activity associated with this phenomenon. When

arriving at the test location, they were given more detailed task instructions. They were told that a fixation point would appear at the screen they would later be watching, and that this point would change between being a simple bar and a cross. When the fixation bar would be shown they should just wait for the fixation cross. When the cross appeared, they were supposed to press either a left or a right button. They were told to wait for the cross and then make a spontaneous choice to press either left or right. Additional importance was placed on them not responding in any stereotyped way or adhering to any specific pattern. An example of such a pattern was given as simply alternating between left and right button presses. Further, they were told that a circle would soon appear after their button press, and that their task then was to estimate the elapsed time between these two events. The perceived time was to be reported immediately by again pressing down the button and holding it down for as long as they thought the interval duration had been. They could then make another button press when the fixation cross again appeared. Importantly, they were not informed about the specific properties of the circles, or the fact that all circles would not be identical. They were neither told anything about the specific durations of the intervals, except that they would be rather short. They were additionally told that the task would be divided into separate blocks and that for each block they would only be allowed to press each button 20 times. Counters in the upper left and right corners of the screen would provide them information on the amount of button presses left for each button. They were told that this provided them with a possibility to check how many presses they had left, but that it was otherwise not something they needed to pay attention to.

2.3.2 Overall procedure

After being given instructions and having completed the safety procedure, the participants were placed in the scanner. They then performed the time estimation task being reported on here. This was divided into two separate sessions, lasting ~ 20 minutes each. Between sessions, scanning was paused and the participants were told that there would now be another session with the exact same task that was equal in length to the previous. After completion of this task, they were told to rest while structural images were acquired, lasting ~ 8 minutes. A second task, not reported on here, was then given to them before they were brought out from the scanner. This task lasted ~ 15 minutes, resulting in a total of ~ 1 hour effective scan time. After scanning, they all completed a short (27-item) questionnaire, which is also not reported on here. They were all then asked about the statistical associations during

the task, specifically whether they noticed that a colour circle would appear more often than a grey circle after pressing one of the buttons and vice versa for the other button.

2.4 Stimulus presentation and data acquisition

2.4.1 Behavioural data

All stimuli were generated and presented, and behavioural data collected, using the Psychophysics Toolbox extensions (Brainard, 1997; Kleiner et al., 2007; Pelli, 1997) implemented in MATLAB R2012b (MathWorks, Natick, MA, USA) running on a Dell Precision T7600. The stimuli were displayed in a mirrored fashion on a 32", 60 Hz (input rate), MR-compatible screen (NNL LCD Monitor®, NordicNeuroLab, Bergen, Norway) placed behind the scanner bore. Participants then watched the screen through a mirror attached to the scanner head-coil. The task was triggered by a pulse from the scanner at the beginning of the fMRI acquisition, synchronising image acquisition with the task. Task responses were given using a set of MR-compatible grips (ResponseGrip®, NordicNeuroLab, Bergen, Norway), one in each hand, with a button that could be pressed using the index fingers.

2.4.2 Imaging data

Brain images were acquired using a Philips Achieva 3 Tesla MRI scanner with an 8-channel Philips SENSE head coil (Philips Medical Systems, Best, The Netherlands). Functional imaging data were collected using a BOLD-sensitive T2*-weighted echo-planar imaging sequence. 34 transversally oriented slices (no gap) were acquired (repetition time (TR), 2 s; echo time (TE), 30 ms; flip-angle, 80°; voxel size, 3 × 3 × 3; field of view (FOV), 240 × 240 mm; interleaved acquisition). The FOV was placed such that both the entire visual and frontal cortex was covered. When larger brain size made this impossible, the most superior part of the cortex (around primary motor/sensory cortices) was left out of the FOV. 6 dummy scans were acquired and discarded before the start of each scan session to avoid T1-saturation effects. The duration of each session varied depending on the speed of that participants button presses from trial to trial, resulting in some variation in the amount of volumes collected. Structural images, consisting of 180 sagittally oriented slices, were acquired using a standard T1-weighted sequence (TR, 8.415 ms; TE, 3.90 ms; flip angle 8°; voxel size 1 × 1 × 1 mm; FOV, 256 × 256 mm).

2.5 Data analysis

All calculations and analyses were performed using Matlab R2011a. If not stated differently, statistics were estimated using the implemented Statistics Toolbox (version 7.5).

2.5.1 Learning model

In estimating trial-wise stimulus-identity predictions, I chose to use the Rescorla-Wagner (RW) model (Rescorla & Wagner, 1972), a simple associative learning model, because of apparent previous success in modelling task-irrelevant learning (den Ouden et al., 2009). For the current purposes, this could be expressed simply as (adapted from den Ouden et al., 2009):

$$\phi_{t+1} = \phi_t + \varepsilon(\lambda_t - \phi)$$

where the stimulus prediction ϕ gets updated on each trial t according to its discrepancy to the actual outcome λ, or in other words, the prediction error. The amount of updating is given by the learning rate ε. This model concerns updating of ϕ in the context of one specific cue where the cue always is identical in terms of saliency. In the original model an additional term is included to denote cue salience. However, this is not relevant for the current study, since the specific cues are always identical.

For each subject here, there were to cues, left and right button press, which could be used to predict whether the subsequent stimulus would be coloured or grey. Predictions were thus calculated separately for each cue, using the above RW equation. Based on the findings of den Ouden et al. (2009), a learning rate of 0.075 was used. For convenience regarding later inferences on the relationship between TB and expectancy/surprise, the outcome of one cue was defined as the opposite to the outcome of the other. For the cue that had a high probabilistic association to colour, colour was defined as a positive outcome in the model. Thus, the predictions associated with this cue were modelled as colour predictions. For the other cue, the predictions were stated in terms of predictions regarding grey. The result of this was two learning curves (fig. 5) with increasing predictions in the early trials, later stabilising with variations around the actual probabilistic contingencies of 0.8.

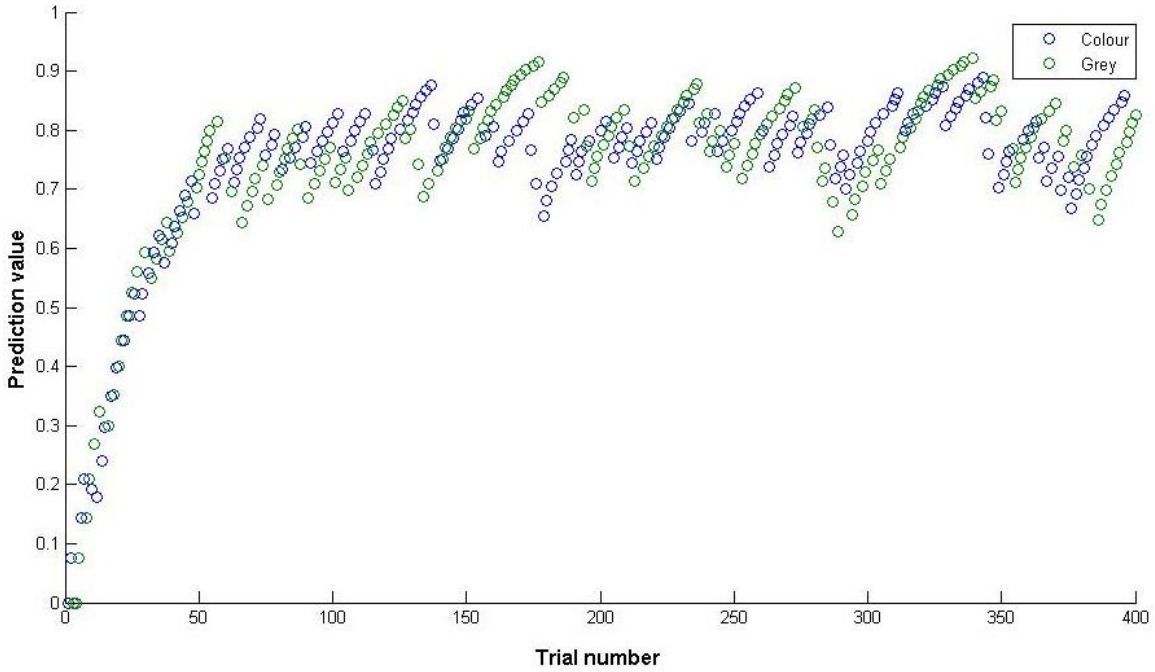


Figure 5. *Example of prediction values calculated for one participant, defined as prediction regarding colour (blue) or grey (green) depending on the cue that was present on the trial.*

2.5.2 Behavioural analysis

From the participants' trial-wise estimations of interval duration, the corresponding TB values were calculated as the estimated interval duration subtracted from the true interval duration. The true interval was defined as the elapsed time from button press to stimulus presentation (onset time), while the estimated interval was defined as the time from button press to button release. In order to make the binding values less dependent on variations in actual length, in slight deviation from previous reports (Poonian & Cunnington, 2013), I further calculated relative TB (rTB) values as the proportion of estimation error to the actual interval length:

$$rTB = \frac{interval_{true} - interval_{estimated}}{interval_{true}}$$

This transformation was done for each response, and the resulting rTB is used as the dependent variable in all subsequent analysis, effectively controlling for TB variation due to actual interval variation.

As a further step in preparing the data for analysis, I tried to identify obvious statistical outliers due to the button being pressed for an abnormally short or long period of time. Since

significant variability in time judgments are expected and considered highly relevant during this task, only data points considered to be extreme outliers were removed from the data set before analysis. This was simply defined as any values more than 4 standard deviations from the mean, resulting in a total of 7 rejected data points or a total data loss of 0.10 %.

Some assumptions regarding the data were then tested. First, a mean rTB score was produced for each participant and subjected to a one-sample t-test to test for relative estimation errors being significantly different from null, and skewed in the expected direction in terms of the classic TB effect. Second, to test whether participants were actually able to perform the task as expected, I estimated the correlation between real and estimated interval for each subject.

In order to avoid a possible non-linear confound with time, the trials containing the initial rise in prediction values, as the associations are learnt, were discarded from further behavioural analysis. To achieve this, I identified the first trial where the prediction had reached the actual probabilistic association of 0.8 and analysed only the following trials for that cue. The same was done for the other cue. Thus, only variations in prediction level after the subject had presumably reached the accurate prediction level was analysed in terms of corresponding rTB variation. The average first trial reaching prediction ≥ 0.8 was 80.74 (SD = 16.31).

In estimating the potential contributions of prediction and prediction error as predictors of variability in TB, I used an approach similar to that of Browning and Harmer (2012), running separate regression models for each subject and using the resulting beta values for later population inference. A general linear model regression analysis was set up individually including regressors of interest and other potentially influencing variables as additional regressors. Of particular interest for the current research questions was the interaction between the trial-specific prediction value and the actual outcome with respect to that prediction (*prediction* × *surprise*). Depending on the outcome, increasing prediction values would represent either increasing surprise or increasing expectancy. Assuming a linear relationship between the predictor and observed response, this interaction term allows testing for the existence of such a relationship, as well as the direction of influence. The components of this interaction (*prediction* and *surprise*) were included as separate regressors to account for any additional main effects. The *surprise* component is thus defined as surprising with respect to the prediction given by the cue. Further, attempting to control for other relevant factors, *time*, *condition*, and *outcome* were included as additional regressors. *Time* was modelled here as a linear increase with trial number. *Condition* was defined as cue type

(associated with colour vs. grey). *Outcome* defined the actual identity of the outcome (colour vs. grey). See supplementary table 1 in the appendix, for an overview of the included regressors with explanations and definitions.

2.5.3 Learning rate optimisation

A weakness in the currently used learning model is that the optimal learning rate during this task in unknown. Since the behavioural and imaging data are statistically independent, it was therefore reasoned that the former could be used as a source for optimising the learning model to be used in analysis of the latter. I therefore ran several regression models for each subject, using incremental learning rates from 0.001 to 0.15 in steps of 0.001 and chose the model with the largest squared t value for the surprise/expectancy interaction (*prediction* × *surprise*) as the one representing the best fit with TB values according to results from the group-level behavioural analysis. However, this produced highly variable results in terms of optimised learning rate (SD = 0.05), as well as some subjects (N = 5) then showing a positive *prediction* × *surprise* interaction (at odds with group behavioural results, see Results section). In order to preserve comparability between behavioural and neural analysis, the initial model with learning rate of 0.075 for all subjects was therefore kept as the best approximation of learning for the group as a whole.

2.5.4 fMRI preprocessing

All preprocessing and analysis of fMRI data were performed using SPM8 (Wellcome Trust Centre for Neuroimaging, London, UK; http://www.fil.ion.ucl.ac.uk/spm) implemented in MATLAB R2011a. To prepare the data for analysis, image coordinates were manually adjusted for each subject so that the origin were set to the anterior commissure (AC) with its axial plane further intersecting the posterior commissure (PC) defining the "Talairach space" (Poldrack et al., 2008), recommended as the starting point for later automated adjustments on the data. The data were then slice-time corrected, adjusting each image slice for differences in acquisition time (Sladky et al., 2011). A realignment procedure was then run to estimate and correct for movement between the acquired volumes (Andersson, Hutton, Ashburner, Turner, & Friston, 2001). The movement parameters were inspected for excessive movement, and none of the subjects were excluded from analysis on this basis. I then coregistered the structural T1 image to the mean image of the already aligned functional EPI's. Successful coregistration was checked by visually inspecting the images match when overlaying the functional images on the structural for each subject. The images were then normalised into

MNI (Montreal Neurological Institute) space by first running a segmentation procedure on the structural image, creating normalisation parameters for that particular subject, before applying those parameters to all of that subjects coregistered functional images. The images were re-written as normalised images using their original voxel size of 3 × 3 × 3. As a last preprocessing step, the normalised functional images were smoothed, using a kernel of 8 × 8 × 8 full-width half-maximum (FWHM) (Mikl et al., 2008), in order to improve signal-to-noise ratio (e.g. Triantafyllou, Hoge, & Wald, 2006).

2.5.5 fMRI analysis

2.5.5.1 Learning analysis

The fMRI-data were analysed as an event-related design using the general linear model (Friston, 1994), creating separate regressors for each trial type which are then convolved with the hemodynamic response function (Friston, Josephs, Rees, & Turner, 1998). The data were high-pass filtered to remove slow signal drifts, using a cut-off set to 128 s. Remaining serial correlations were modelled using a first-order autoregressive model (Friston et al., 2002).

For each trial in the experiment, there were 2 relevant variables (*cue* and *outcome*) with 2 levels each (colour-predictive/grey-predictive and colour/grey), which resulted in 4 possible ways of defining a given trial, and thus initially 4 different regressors in the design matrix. These regressors were defined according to the onset times of the outcomes on each trial. For the current research question, I was however only interested in neural activity associated with the hypothesised stimulus-learning previously tested for using the behavioural data. Following this investigation, I used the set of prediction values previously described with predictions defined according to the cue, resulting in two sets of approximately parallel learning curves, one denoting colour prediction and the other grey prediction. These values were then used as input in the analysis as parametric modulators (PM) for their respective regressors. Specifically, the outcomes of the button press (cue) that predicted colour was modulated by trial-specific levels of colour prediction and the outcomes of the other cue was modulated by grey prediction levels, defined separately for each cue-outcome combination. In this way, the PM of each regressor would capture activity associated with either increasing levels of expectancy or increasing levels of surprise for the outcome given the cue (fig. 6). This was then analysed further in a second-level analysis. Each PM was further expanded by a 2^{nd} order polynominal to account for some additional non-linearity, making the model less

dependent on the learning rate of 0.075 as well as between-subject variability in the shape of the learning curve (den Ouden et al., 2009). Using simple condition-specific t-contrasts of the PM and PM expansion from each subject, I performed a group-level analysis using a 2 (*cue*: colour-predictive vs. grey-predictive) × 2 (*outcome*: colour vs. grey) repeated-measures ANOVA with an added subject-factor. With the PM's as input, this was effectively testing for the 3-way interaction of *prediction* × *cue* × *outcome*, showing either activity related to surprise or expectancy depending on the direction of the interaction.

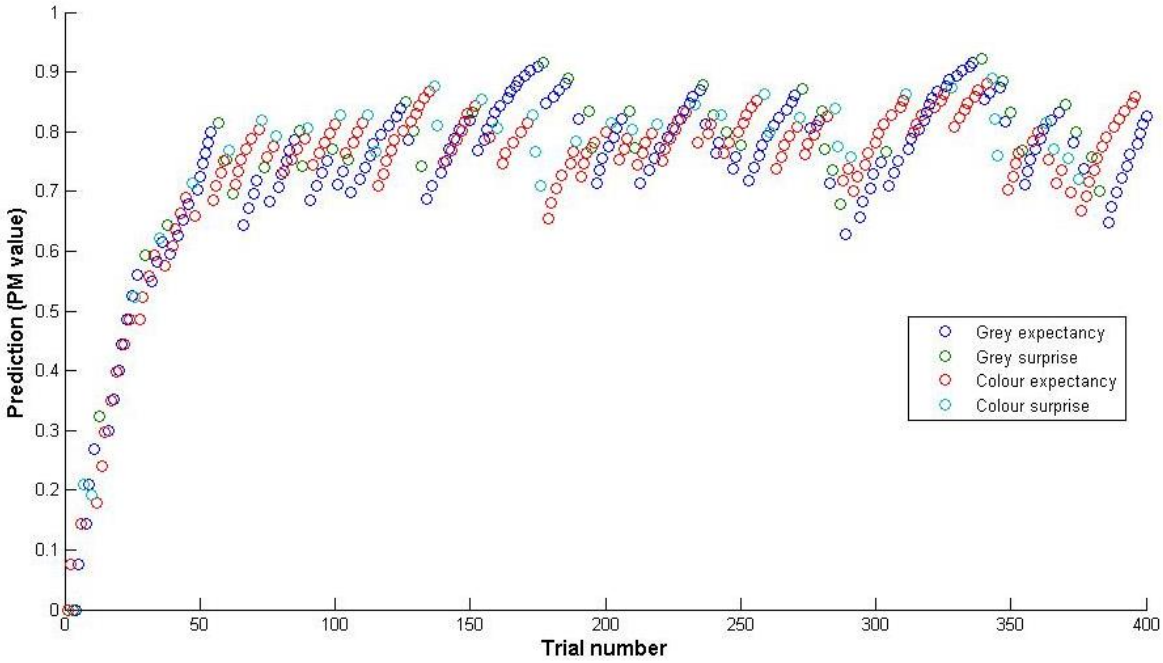


Figure 6. *The predictions defined as PMs according to the four different regressors. The PMs capture activity associated with variation in expectancy and surprise for each outcome.*

In order to maximise the chance of finding any true activation in the data, I used anatomically restricted search-volumes based on previous knowledge. Following den Ouden et al. (2009), I expected prediction error related activity in colour-sensitive areas in the occipital cortex as well as putamen and possibly also caudate nucleus. For a learning response independent of outcome, the right inferior frontal gyrus was also included as a candidate area. The Talairach Daemon atlas (Lancaster et al., 2000) within the WFU PickAtlas toolbox (version 2.4) (Maldjian, Laurienti, Kraft, & Burdette, 2003) in SPM was used to create a mask containing these regions. The entire occipital cortex was included to account for multiple colour-sensitive areas (Beauchamp et al., 1999) as well as other perceived differences between the stimuli (e.g. perceived luminance) that was not controlled for in the task. Using

activation maps thresholded by $p < 0.005$ with a 10 voxel extent (Lieberman & Cunningham, 2009), I then performed a small-volume correction (SVC) (Poldrack, 2007) using this mask for the interaction contrasts. For each subject, a t contrast was also defined for the main effect of the PMs. This was then subjected to a one-sample t test to test for a main effect of prediction regardless of outcome. Age, gender, and handedness were entered as covariates. The same SVC was then also performed on this group activation map.

2.5.5.2 Comparative analysis and re-analysis based on SUIT normalisation

Since a recent study has investigated neural correlates of TB resulting from a direct estimation task (Kühn et al., 2013), this was used as an opportunity to compare results from the current data set. By comparing TB-related activation during the current task with the previous findings, this would give an indication to both the validity of the task as reflecting true TB as well as the validity of the fMRI design in characterising the underlying neural signal. I therefore proceeded by using the previously described preprocessed images in a new analysis. For this analysis, only 2 regressors were used, one for each button press type (right vs. left), and all onset times were redefined to the button press onset, following Kühn et al. (2013). The rTB values were now entered as PMs for their respective regressors. Otherwise, all of the first-level settings were identical to the previously described analysis. Group-level activation was tested for using a one-sample t-test on individually defined contrasts specifying activation associated with increase and decrease in rTB. Age, gender and handedness were included as covariates. A mask was defined for supplementary motor area (SMA) using the WFU PickAtlas to test for a replication of the results of Kühn et al. (2013).

Because I found indications of upper brain-stem activity during this comparative analysis (see Results), I wanted to run an extra procedure to better characterise its probable anatomical location. Because the regular normalisation method used by SPM has been found to do a poor job of normalising the cerebellum and brain-stem areas (Diedrichsen, 2006), I normalised the previously realigned and coregistered images again using the SUIT toolbox in SPM, which is based on a spatially unbiased template of the cerebellum and brain-stem (Diedrichsen, 2006; Diedrichsen, Balsters, Flavell, Cussans, & Ramnani, 2009; Diedrichsen et al., 2011). Because the level of smoothing can affect the ability to localise the inferred activation (e.g. Geissler et al., 2005), I also used a reduced smoothing kernel of $6 \times 6 \times 6$ FWHM before analysing these images again using the above described design.

3 Results

When asked whether they noticed that the colour stimulus showed up more frequently after pressing one of the buttons and that grey showed up more often after the other button press, 3 participants said they noticed it, 2 said that they had wondered about it at some point during the experiment, and the remaining 12 said they had not noticed it at all.

3.1 Behaviour

The subjects' average rTB values show that the intervals are judged as significantly shorter than the real interval lengths at the group-level ($t(16) = 3.47$, $p = 3.17 \times 10^{-3}$) with a group-average rTB value of 0.16 (SE = 0.05) in line with the TB effect. However, it should also be noted that 3 participants actually had negative average rTB values, on average judging the intervals as longer than their actual lengths. The correlations between estimated and real intervals were positive and significant for all participants, indicating at least decent performance of the task. However, performance in this respect was also highly variable (r between 0.20 ($p = 6.16 \times 10^{-5}$) and 0.75 ($p = 5.10 \times 10^{-73}$); $mean_r = 0.53$; $SD_r = 0.16$).

From analysing the betas from the individual regression analyses, the interaction between prediction value and surprising outcome was found to be significant across the group (*prediction* × *surprise*: $t(16) = -2.67$, $p_{uncorr} = 0.02$). The nature of this interaction was negative, such that higher prediction values are predictive of lower TB values when the outcome is surprising with respect to that prediction. Additionally, whether the outcome was surprising was found to be an additionally significant predictor across subjects (*surprise*: $t(16) = 2.65$, $p_{uncorr} = 0.02$). Also, time was found to be a significant predictor (*time*; $t(16) = -3.11$, $p_{uncorr} = 0.007$). Accounting for multiple comparisons across the 6 predictors of interest using a positive false discovery rate (pFDR) estimation (Storey, 2002) yields acceptable values (< 0.05) for all above p values.

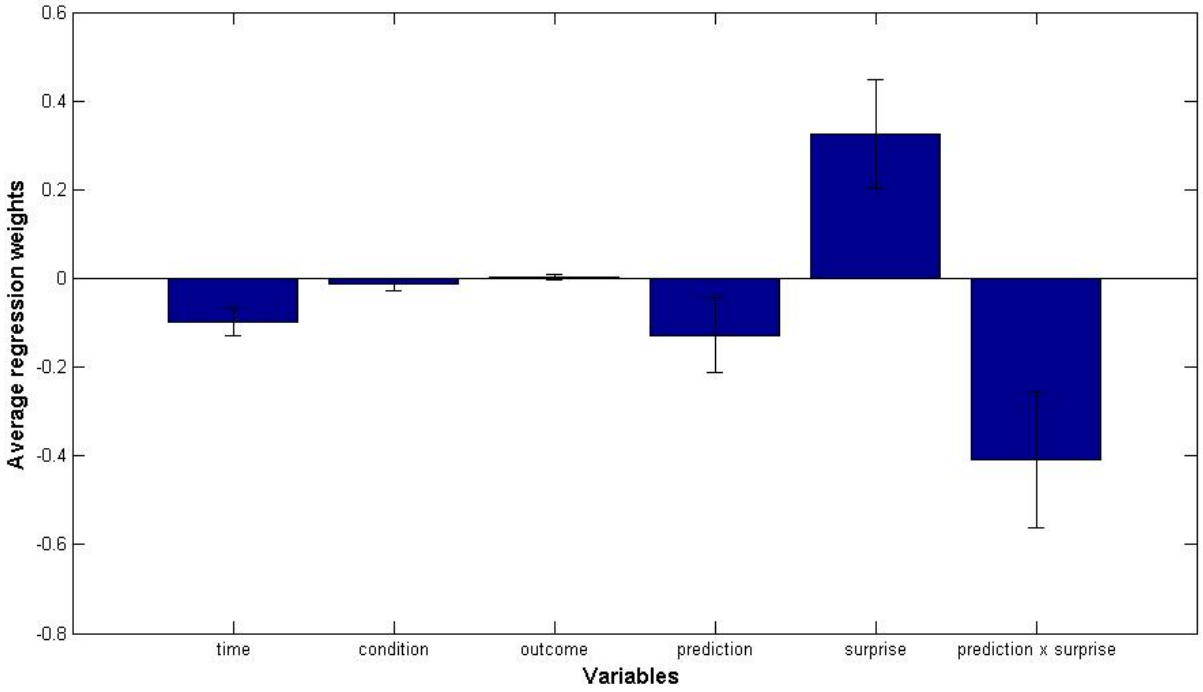


Figure 7. *Bar plots showing average beta weights across the group. Error bars indicate standard errors of the mean.*

3.2 fMRI

3.2.1 Learning analysis

No significant activation was found for the 3-way interaction *prediction* × *cue* × *outcome* using the anatomically defined mask for SVC ($p_{FWE} < 0.05$). A whole-brain search was then performed, checking for significant activation surviving a family-wise error (FWE) rate of 0.05, when thresholded at $p < 0.005$ and $k > 10$ (Lieberman & Cunningham, 2009). Still, no activation was identified, not even when setting the significance level to $p_{uncorr} = 0.001$ ($k > 20$). I did an additional check of this finding by taking individually defined t contrasts for surprise compared with expectancy and running a one-sample t test for the group. This resulted in a similar null finding (for $p_{uncorr} = 0.001$, $k > 20$). Additionally, no activation was found for the 2-way interactions *prediction* × *cue* or *prediction* × *outcome* using SVC ($p_{FWE} < 0.05$) or whole brain FWE of 0.05. The one-sample t test on the main effect of *prediction* revealed no activation either, again using the same two approaches to correct for multiple comparisons.

3.2.2 Comparative analysis

The SMA ROI showed activation associated with neither increase nor decrease in rTBs. I therefore checked the same contrasts at the whole-brain level thresholded at $p < 0.005$ and $k > 10$. For the positive correlation with rTBs, still no activity was found to be significant. However, for the negative correlation, that is, activation associated with decreasing rTBs (longer estimations, when corrected for actual interval length), two clusters were identified as significant at the cluster level ($p_{FWE} = 9.74 \times 10^{-3}$ and $p_{FWE} = 2.86 \times 10^{-5}$). The largest cluster ($k = 391$) was located in the precuneus (peak MNI -18, -52, 31) and the other cluster ($k = 170$) with peak-activation in the left upper brain-stem (MNI -15, -22, -11) extending further to the left hippocampus (fig. 8). Acknowledging the fact that the extent of the second cluster is anatomically improbable, and at best contains several overlapping clusters, it was still considered worthwhile to take a closer look at the anatomical location of the peak, even though the peak itself was not significant after correcting for multiple comparison (pFWE = 0.37, Z = 4.27).

Although no activity survived correction for multiple comparisons for the positive correlation with rTBs, the uncorrected t map was indicative of a potentially activated network (see supplementary fig. 1, Appendix).

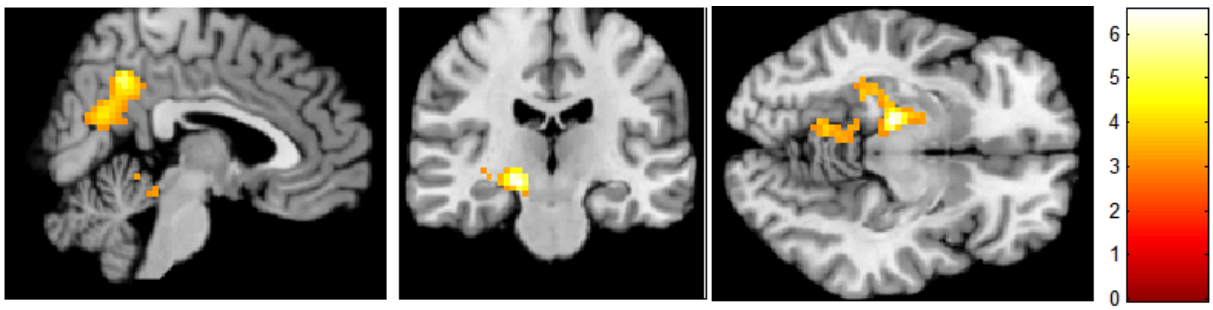


Figure 8. *The two significant activation clusters identified for a decrease in rTBs. The activation map is overlaid on a standard MNI template (colin27). Colour bar denotes t values.* $p_{uncorr} < 0.005$ *for illustration purpose.*

3.2.2.1 Re-analysis after SUIT normalisation

After re-analysing the data after improved normalisation and less smoothing, the peak was more specifically located at MNI -16, -25, -6 ($p_{uncorr} = 1.41 \times 10^{-5}$, Z = 4.19) and activation was found to overlap partly with the substantia nigra (fig. 9) as defined using WFU PickAtlas.

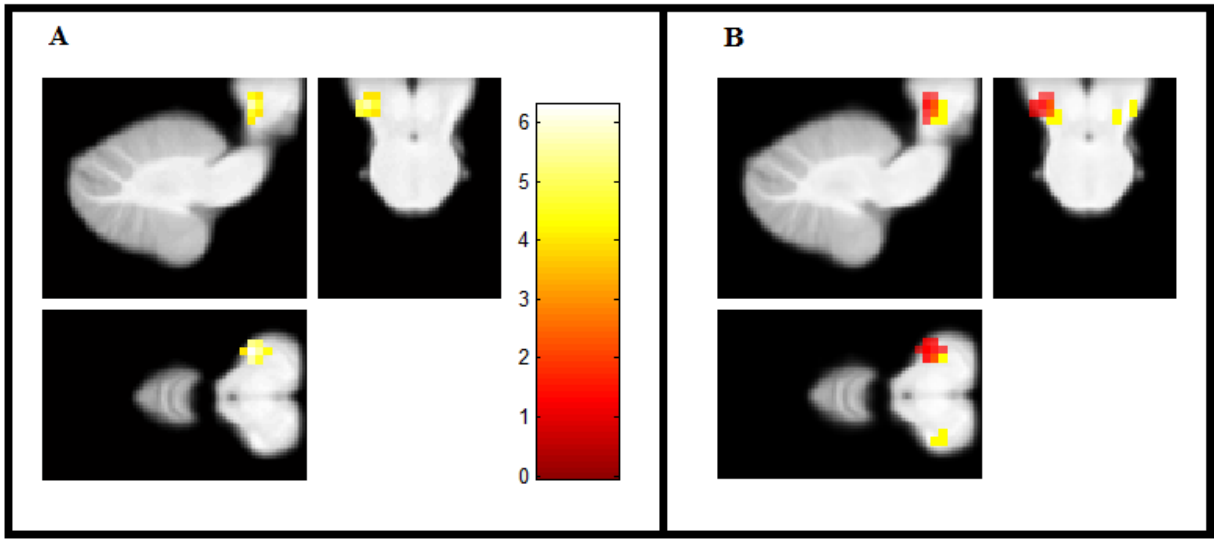


Figure 9. *(A) Brain-stem activation (p_{uncorr} = 0.001) after re-analysing the data using the SUIT normalisation procedure and (B) partly overlap between activation cluster (red) and atlas definition of substantia nigra (yellow). Activation and atlas definition are overlaid on a group-averaged T1 image resulting from isolating and normalising the brain-stem and cerebellum in SUIT.*

4 Discussion

4.1 Predictive learning and TB

Contrasting with earlier claims (Desantis et al., 2012), I here provide evidence that TB is indeed affected by stimulus-specific predictions. The outcomes were displayed with equal frequencies across blocks during the experiment, so the only way for participants to predict specific outcomes (and get surprised by violations) was to use the action as cue. The comparison between the learning model and behavioural data further suggests that the surprising stimuli result in decreasing TB with increasing predictions regarding those stimuli. In other words, increased predictability results in increased binding if the stimulus matches the prediction and decreased binding if the stimulus violates the prediction. This is as hypothesised and matches an explanation of TB as resulting from a comparison between internal sensory predictions and externally generated sensations. There was also a trend in the data towards predictability having an effect on TB regardless of outcome, but this was not statistically significant. This could well be a result of the limited sample size used here, since other studies have shown binding when the probability of an effect is high regardless of outcome (Engbert & Wohlschläger, 2007).

In spite of the indications from the behavioural data, I did not find any evidence of neural activation associated with expectancy or surprise during performance of the task. This could be due to lack of power in the fMRI design or a bad model fit due to the chosen learning rate. This last point is however not a sufficient explanation since it was found that this particular learning model had an influence on the TB measures. If this inference is correct, there should also be neural activity associated with the model. Because the attempt of individually optimising learning curves yielded non-consistent results, the behavioural results was considered being non-consistent across subjects. This indicated that some participants did not show any effect of expectancy/surprise on TB or that some might even show a different pattern than the majority of the group. Because of this, it was not considered possible to estimate individually tailored learning curves, and the initial learning model, showing a group-level fit, was therefore chosen. As a result of these rather weak results it might be that the effect in the behavioural data was too small to be picked up in the imaging data.

The finding that surprising outcomes exert an effect on binding regardless of the specified prediction value, further suggests that the learning model has not sufficiently accounted for all prediction variability. Some of this lack of model fit could perhaps be

attributed to the non-optimal learning rate that was chosen or it might well be that the model is not sophisticated enough to account for all variability. As such, there seems to be potential for future research in working towards a better model of surprise-induced learning in TB. Additionally, the main effect of surprising outcomes is actually related to binding in a positive way. It contributes to an increase in binding. This is unexpected, since a match of outcome and sensory prediction has been thought to increase binding, and not the other way around. Similarly, the main effect of the prediction, even though non-significant, indicates a negative influence of prediction on TB. The mechanisms underlying the resulting interaction between prediction and surprising outcome may therefore also be different than initially hypothesised. It might even be that stimulus-specific surprise affects binding differently than surprise regarding the actual occurrence of any stimulus. This distinction might be the subject of future studies.

In line with recent evidence, the findings here support a change in TB over time (Cravo, Haddad, Claessens, & Baldo, 2013). However, in contrast with those findings, the current data support a decrease in TB, or a general lengthening of judgments over time. This could be due to differences in measuring TB.

4.2 Neural correlates of TB

The failure to replicate the findings of Kühn et al. (2013) regarding SMA activation associated with TB and the identification of new neural patterns, places doubt on the assumption that we have been studying the same phenomenon. Apart from the components in the task, the only difference between our approaches was the method used to measure TB. This should therefore be given some further thought. Although both studies used a direct estimation procedure, Kühn et al. (2013) used the more common method of having the subjects report the interval using a scale. As previously noted, this necessarily involves a transformation of the experienced duration to some sort of number system in order to convey the experience, which at least intuitively could seem like a more complex task and thus possibly introduce more uncertainty, or even bias, in the measurements. The second difference between the two approaches pertains to the length and variation in the intervals that was used. I currently used continuously varying intervals, whereas Kühn et al. (2013) used only three distinct intervals. Further, the current intervals varied between 500 and 1500 ms, compared with intervals of 200, 300 or 400 ms in the previous study. This is of special importance since varying degrees of binding has been reported at different interval lengths under certain estimation methods (Humphreys & Buehner, 2009). In combination with the

previously described behavioural results contrasting with previous findings, this suggests that the methods used to measure TB may not measure the same phenomena. Therefore, future studies on phenomenological implications of these methodological differences are highly encouraged.

The exact location of the brain-stem activation identified should be interpreted with extreme caution because of limited spatial resolution in the images as well as further smoothing of the signal during preprocessing. However, the peak activation appears close to the left substantia nigra. Since the involvement of this structure has been reported during a simple interval reproduction task similar to the one used here (Jahanshahi, Jones, Dirnberger, & Frith, 2006), it could be hypothesised that it plays a role beyond simple reproduction. The activation reported here does not describe reproduction per se, but instead the direction of reproduction error during the reproduction task. The brain-stem activity correlates negatively with shorter estimates, corrected for actual length. That is, the longer the temporal estimates relative to the actual length, the more activation there is. It seems unlikely that this activation should be related to a decrease in sense of agency. It is therefore assumed that this must be related to another aspect of the task, like the direction of reproduction error.

Activation of the precuneus has previously been reported to be associated with attribution of action to another person compared with attribution to oneself (Farrer & Frith, 2002). It has further been implied in the neural network associated with self-consciousness (Cavanna & Trimble, 2006). Perhaps most interestingly, this is a central node in the so-called 'default-mode network' (Fransson & Marrelec, 2008; Raichle et al., 2001; Utevsky, Smith, & Huettel, 2014) which consistently shows deactivation during demanding tasks. If it is assumed that variations in rTB reflect variations in agency, it could certainly make sense that a decrease in agency could be linked to a task-deactivation node. If the non-significant activation pattern for the positive rTB contrast is used as a further indication (supplementary fig. 1, appendix), one could speculate that the positive association might represent activation of a network associated with attention or otherwise be a result of increased task-demand. Consistent with the speculation that this region is somehow involved in agency is the finding of abnormal connectivity in this area in schizophrenia (Bluhm et al., 2007) as well as failures of deactivation in the same patient group (Salgado-Pineda et al., 2011) and correlation with these patients' positive symptoms (Garrity et al., 2007).

A study on patients with Parkinson's disease also showed failure of task-related precuneus deactivation which was restored with use of dopaminergic medication (Delaveau et al., 2010).

Since no control condition was included in the current study, there can be made no inference as to whether any of the observed activity is due to voluntary action specifically. It is therefore possible that none of the clusters might be specific to voluntary action. It is also possible that one of them is, whereas the other is due to some other aspect of the task. Because of the central function of the precuneus in the default mode network, I speculate that this activation reflects participants being less attentive to or involved in the task. If this is true, there might be an increased 'binding baseline' when attentional resources are directed at the task compared to when they are not. It would thus have been interesting to see a study investigating a potential modulating role of attention on binding.

4.3 Limitations

As previously mentioned, there is an important limitation in the current study in the exclusion of a control condition for non-specific voluntary action effects. Thus, the occurrence of a true TB effect can not be validated for the current task.

The regression model that was used in the behavioural analysis assumed a linear relationship between predictors and outcome. This might of course be a false assumption, implying that the current division between interacting and specific contributions of prediction level and outcome might be a false dichotomy. Future studies might thus find evidence of more specific prediction-outcome relations for TB. The results are, nonetheless, indicative of an interaction between the participants stimulus-specific expectations and the nature of the outcome regarding that particular expectation.

Because participants were given a cue (fixation cross) when they were allowed to press the button, their button press could be seen as partly reactive, not giving them the freedom to press at a time of their own choosing. They were, however, carefully instructed not to make a choice of which button to press before the fixation cross appeared, thus retaining a spontaneous choice (for all but the last trial(s) in each block). Stated differently, they were given the choice of which button to press, but not when to press it. It is likely that this is a somewhat different phenomenon than freely choosing when to act and could contribute to the current results contrasting with earlier findings.

Further, because left and right counters were used in the task, it can be argued that the participants' attention could be distracted away from the fixation point also during the task. However, it is unlikely that this would have a significant impact, because performance of the task necessitates attention towards the location where the stimulus will appear.

Since the intervals are all estimated retrospectively, we can't know whether variations in TB are representing real perceptual errors or if it is due to memory and recollection processes. More fundamentally, we don't know whether time judgments reflect the subjective experience of time.

4.4 Future directions

Based on the current findings, it could be argued for a central role for prediction error in affecting TB. The stimuli used here differed only in a rather minor way, by varying one simple feature. Even the highest value of prediction error obtained here could thus be regarded as quite small when considering that most of the prediction was always fulfilled (the stimulus was always a circle with the same pattern). It would therefore be interesting if future investigations would take into account such differences in prediction error. If a study is designed where the stimulus identity differs quantitatively with respect to how well it matches the prediction, such effects could be tested for.

The failure to identify a neural prediction error response when such prediction error seems to affect the task, could provide a case for future work to investigate how TB could relate to neural hierarchical predictive coding, especially at lower levels involving basic perceptual processes. Since it has been suggested that TB reflects a 'preconscious' sense of agency that could be dissociated from the type of agency that one is consciously aware of and capable of reporting verbally, it would be particularly interesting to know if these processes are fundamentally different or whether they relate to similar principles operating at different hierarchical levels. Theoretical work on how predictive coding relates to agency and sensory attenuation has already emerged (Brown, Adams, Parees, Edwards, & Friston, 2013).

As already mentioned, more work is encouraged on characterising the impact of methodological differences in measuring TB, especially how this relates to subjective experiences of motor control and sense of agency. Also, attempts should be made on replicating and extending previous findings regarding neural activity associated with binding, using different TB measures.

4.5 Concluding remarks

The current study provides evidence that predictions of stimulus-specific features are learnt during action-effect interval estimation, and that these predictions do affect temporal judgments, even when completely task-irrelevant. It is concluded that the predictions affect temporal binding through an interaction with their respective outcomes, that is, through

prediction error. Specifically, such an increased prediction error leads to a decrease in binding. However, because the main effect of surprise here is predictive of an increase in TB, any conclusions of a general effect of prediction error can not be drawn from the current study. Even though no independent contribution of prediction level was found here, an independent role for stimulus-prediction regardless of outcome can not be ruled out.

There was not found any evidence of a neural prediction or prediction error response during this task, which leaves the question of how such learning is implemented in the brain still open. However, activation of the precuneus was found to be associated with decreased binding. I speculate that this might represent 'default mode' activation and that it might be a result of less resources being allocated towards processing of task-demands. Future studies are therefore encouraged to investigate a possible modulating effect of attention on binding.

5 References

Aarts, H., Bijleveld, E., Custers, R., Dogge, M., Deelder, M., Schutter, D., & van Haren, N. E. M. (2012). Positive priming and intentional binding: Eye-blink rate predicts reward information effects on the sense of agency. *Social Neuroscience, 7*(1), 105-112. doi: 10.1080/17470919.2011.590602

Aarts, H., Custers, R., & Wegner, D. M. (2005). On the inference of personal authorship: Enhancing experienced agency by priming effect information. *Consciousness and Cognition, 14*(3), 439-458. doi: http://dx.doi.org/10.1016/j.concog.2004.11.001

Adams, R. A., Shipp, S., & Friston, K. J. (2013). Predictions not commands: active inference in the motor system. *Brain Structure and Function, 218*(3), 611-643. doi: 10.1007/s00429-012-0475-5

Andersson, J. L., Hutton, C., Ashburner, J., Turner, R., & Friston, K. (2001). Modeling geometric deformations in EPI time series. *NeuroImage, 13*(5), 903-919.

Bayer, H. M., & Glimcher, P. W. (2005). Midbrain Dopamine Neurons Encode a Quantitative Reward Prediction Error Signal. *Neuron, 47*(1), 129-141. doi: http://dx.doi.org/10.1016/j.neuron.2005.05.020

Bays, P. M., Flanagan, J. R., & Wolpert, D. M. (2006). Attenuation of Self-Generated Tactile Sensations Is Predictive, not Postdictive. *PLoS Biol, 4*(2), e28. doi: 10.1371/journal.pbio.0040028

Beauchamp, M. S., Haxby, J. V., Jennings, J. E., & DeYoe, E. A. (1999). An fMRI Version of the Farnsworth–Munsell 100-Hue Test Reveals Multiple Color-selective Areas in Human Ventral Occipitotemporal Cortex. *Cerebral Cortex, 9*(3), 257-263. doi: 10.1093/cercor/9.3.257

Blakemore, S.-J. (2009). How We Recognize Our Own Actions. In N. Murphy, G. R. Ellis & T. O'Connor (Eds.), *Downward Causation and the Neurobiology of Free Will* (pp. 145-151): Springer Berlin Heidelberg.

Blakemore, S.-J., Wolpert, D., & Frith, C. (2000). Why can't you tickle yourself? *NeuroReport, 11*(11).

Blakemore, S.-J., Wolpert, D. M., & Frith, C. D. (2002). Abnormalities in the awareness of action. *Trends in cognitive sciences, 6*(6), 237-242. doi: http://dx.doi.org/10.1016/S1364-6613(02)01907-1

Bluhm, R. L., Miller, J., Lanius, R. A., Osuch, E. A., Boksman, K., Neufeld, R., . . . Williamson, P. (2007). Spontaneous Low-Frequency Fluctuations in the BOLD Signal

in Schizophrenic Patients: Anomalies in the Default Network. *Schizophrenia Bulletin, 33*(4), 1004-1012. doi: 10.1093/schbul/sbm052

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision, 10*(4), 433-436. doi: doi:10.1163/156856897X00357

Brown, H., Adams, R., Parees, I., Edwards, M., & Friston, K. J. (2013). Active inference, sensory attenuation and illusions. *Cognitive Processing, 14*(4), 411-427. doi: 10.1007/s10339-013-0571-3

Browning, M., & Harmer, C. J. (2012). Expectancy and surprise predict neural and behavioral measures of attention to threatening stimuli. *NeuroImage, 59*(2), 1942-1948. doi: http://dx.doi.org/10.1016/j.neuroimage.2011.09.007

Buehner, M. J. (2012). Understanding the Past, Predicting the Future: Causation, Not Intentional Action, Is the Root of Temporal Binding. *Psychological Science, 23*(12), 1490-1497. doi: 10.1177/0956797612444612

Buehner, M. J., & Humphreys, G. R. (2009). Causal Binding of Actions to Their Effects. *Psychological Science, 20*(10), 1221-1228. doi: 10.1111/j.1467-9280.2009.02435.x

Cardoso-Leite, P., Mamassian, P., Schütz-Bosbach, S., & Waszak, F. (2010). A New Look at Sensory Attenuation: Action-Effect Anticipation Affects Sensitivity, Not Response Bias. *Psychological Science, 21*(12), 1740-1745. doi: 10.1177/0956797610389187

Cavanna, A. E., & Trimble, M. R. (2006). The precuneus: a review of its functional anatomy and behavioural correlates. *Brain, 129*(3), 564-583. doi: 10.1093/brain/awl004

Choi, H., & Scholl, B. J. (2006). Perceiving causality after the fact: Postdiction in the temporal dynamics of causal perception. *Perception, 35*(3), 385-399. doi: 10.1068/p5462

Conant, R. C., & Ross Ashby, W. (1970). Every good regulator of a system must be a model of that system †. *International Journal of Systems Science, 1*(2), 89-97. doi: 10.1080/00207727008920220

Cravo, A. M., Claessens, P. M. E., & Baldo, M. V. C. (2009). Voluntary action and causality in temporal binding. *Experimental Brain Research, 199*(1), 95-99. doi: 10.1007/s00221-009-1969-0

Cravo, A. M., Haddad, H., Claessens, P. M. E., & Baldo, M. V. C. (2013). Bias and learning in temporal binding: Intervals between actions and outcomes are compressed by prior bias. *Consciousness and Cognition, 22*(4), 1174-1180. doi: http://dx.doi.org/10.1016/j.concog.2013.08.001

de Vignemont, F., & Fourneret, P. (2004). The sense of agency: A philosophical and empirical review of the “Who” system. *Consciousness and Cognition, 13*(1), 1-19. doi: http://dx.doi.org/10.1016/S1053-8100(03)00022-9

Delaveau, P., Salgado-Pineda, P., Fossati, P., Witjas, T., Azulay, J.-P., & Blin, O. (2010). Dopaminergic modulation of the default mode network in Parkinson's disease. *European Neuropsychopharmacology, 20*(11), 784-792. doi: http://dx.doi.org/10.1016/j.euroneuro.2010.07.001

Demanet, J., Muhle-Karbe, P. S., Lynn, M. T., Blotenberg, I., & Brass, M. (2013). Power to the will: How exerting physical effort boosts the sense of agency. *Cognition, 129*(3), 574-578. doi: http://dx.doi.org/10.1016/j.cognition.2013.08.020

den Ouden, H. E. M., Friston, K. J., Daw, N. D., McIntosh, A. R., & Stephan, K. E. (2009). A Dual Role for Prediction Error in Associative Learning. *Cerebral Cortex, 19*(5), 1175-1185. doi: 10.1093/cercor/bhn161

Desantis, A., Hughes, G., & Waszak, F. (2012). Intentional Binding Is Driven by the Mere Presence of an Action and Not by Motor Prediction. *PLoS ONE, 7*(1), e29557. doi: 10.1371/journal.pone.0029557

Diedrichsen, J. (2006). A spatially unbiased atlas template of the human cerebellum. *NeuroImage, 33*(1), 127-138. doi: http://dx.doi.org/10.1016/j.neuroimage.2006.05.056

Diedrichsen, J., Balsters, J. H., Flavell, J., Cussans, E., & Ramnani, N. (2009). A probabilistic MR atlas of the human cerebellum. *NeuroImage, 46*(1), 39-46. doi: http://dx.doi.org/10.1016/j.neuroimage.2009.01.045

Diedrichsen, J., Maderwald, S., Küper, M., Thürling, M., Rabe, K., Gizewski, E. R., . . . Timmann, D. (2011). Imaging the deep cerebellar nuclei: A probabilistic atlas and normalization procedure. *NeuroImage, 54*(3), 1786-1794. doi: http://dx.doi.org/10.1016/j.neuroimage.2010.10.035

Eagleman, D. M., & Holcombe, A. O. (2002). Causality and the perception of time. *Trends in cognitive sciences, 6*(8), 323-325. doi: http://dx.doi.org/10.1016/S1364-6613(02)01945-9

Eagleman, D. M., & Sejnowski, T. J. (2000). Motion Integration and Postdiction in Visual Awareness. *Science, 287*(5460), 2036-2038. doi: 10.1126/science.287.5460.2036

Eagleman, D. M., & Sejnowski, T. J. (2003). The line-motion illusion can be reversed by motion signals after the line disappears. *Perception, 32*(8), 963-968. doi: 10.1068/p3314a

Engbert, K., & Wohlschläger, A. (2007). Intentions and expectations in temporal binding. *Consciousness and Cognition, 16*(2), 255-264. doi: http://dx.doi.org/10.1016/j.concog.2006.09.010

Engbert, K., Wohlschläger, A., & Haggard, P. (2008). Who is causing what? The sense of agency is relational and efferent-triggered. *Cognition, 107*(2), 693-704. doi: http://dx.doi.org/10.1016/j.cognition.2007.07.021

Engbert, K., Wohlschläger, A., Thomas, R., & Haggard, P. (2007). Agency, subjective time, and other minds. *Journal of Experimental Psychology: Human Perception and Performance, 33*(6), 1261-1268. doi: 10.1037/0096-1523.33.6.1261

Farrer, C., & Frith, C. D. (2002). Experiencing Oneself vs Another Person as Being the Cause of an Action: The Neural Correlates of the Experience of Agency. *NeuroImage, 15*(3), 596-603. doi: http://dx.doi.org/10.1006/nimg.2001.1009

Fransson, P., & Marrelec, G. (2008). The precuneus/posterior cingulate cortex plays a pivotal role in the default mode network: Evidence from a partial correlation network analysis. *NeuroImage, 42*(3), 1178-1184. doi: http://dx.doi.org/10.1016/j.neuroimage.2008.05.059

Friston, K. J. (1994). Functional and effective connectivity in neuroimaging: A synthesis. *Human Brain Mapping, 2*(1-2), 56-78. doi: 10.1002/hbm.460020107

Friston, K. J. (2010). The free-energy principle: a unified brain theory? [10.1038/nrn2787]. *Nat Rev Neurosci, 11*(2), 127-138. doi: http://www.nature.com/nrn/journal/v11/n2/suppinfo/nrn2787_S1.html

Friston, K. J., Daunizeau, J., & Kiebel, S. J. (2009). Reinforcement Learning or Active Inference? *PLoS ONE, 4*(7), e6421. doi: 10.1371/journal.pone.0006421

Friston, K. J., Daunizeau, J., Kilner, J., & Kiebel, S. (2010). Action and behavior: a free-energy formulation. *Biological Cybernetics, 102*(3), 227-260. doi: 10.1007/s00422-010-0364-z

Friston, K. J., Glaser, D. E., Henson, R. N. A., Kiebel, S., Phillips, C., & Ashburner, J. (2002). Classical and Bayesian Inference in Neuroimaging: Applications. *NeuroImage, 16*(2), 484-512. doi: http://dx.doi.org/10.1006/nimg.2002.1091

Friston, K. J., Josephs, O., Rees, G., & Turner, R. (1998). Nonlinear event-related responses in fMRI. *Magnetic Resonance in Medicine, 39*(1), 41-52. doi: 10.1002/mrm.1910390109

Friston, K. J., Kilner, J., & Harrison, L. (2006). A free energy principle for the brain. *Journal of Physiology-Paris, 100*(1–3), 70-87. doi: http://dx.doi.org/10.1016/j.jphysparis.2006.10.001

Friston, K. J., Mattout, J., & Kilner, J. (2011). Action understanding and active inference. *Biological Cybernetics, 104*(1-2), 137-160. doi: 10.1007/s00422-011-0424-z

Friston, K. J., Shiner, T., FitzGerald, T., Galea, J. M., Adams, R., Brown, H., . . . Bestmann, S. (2012). Dopamine, Affordance and Active Inference. *PLoS Comput Biol, 8*(1), e1002327. doi: 10.1371/journal.pcbi.1002327

Frith, C. D. (2012). Explaining delusions of control: The comparator model 20 years on. *Consciousness and Cognition, 21*(1), 52-54. doi: http://dx.doi.org/10.1016/j.concog.2011.06.010

Galea, J. M., Bestmann, S., Beigi, M., Jahanshahi, M., & Rothwell, J. C. (2012). Action Reprogramming in Parkinson's Disease: Response to Prediction Error Is Modulated by Levels of Dopamine. *The Journal of Neuroscience, 32*(2), 542-550. doi: 10.1523/jneurosci.3621-11.2012

Gallistel, C. R. (2009). *The importance of proving the null.* 116 doi:10.1037/a0015251, American Psychological Association, US.

Garrity, A., Pearlson, G., McKiernan, K., Lloyd, D., Kiehl, K., & Calhoun, V. (2007). Aberrant "default mode" functional connectivity in schizophrenia. *American journal of psychiatry, 164*(3), 450-457.

Geissler, A., Lanzenberger, R., Barth, M., Tahamtan, A. R., Milakara, D., Gartus, A., & Beisteiner, R. (2005). Influence of fMRI smoothing procedures on replicability of fine scale motor localization. *NeuroImage, 24*(2), 323-331. doi: http://dx.doi.org/10.1016/j.neuroimage.2004.08.042

Glimcher, P. W. (2011). Understanding dopamine and reinforcement learning: The dopamine reward prediction error hypothesis. *Proceedings of the National Academy of Sciences, 108*(Supplement 3), 15647-15654. doi: 10.1073/pnas.1014269108

Haggard, P. (2005). Conscious intention and motor cognition. *Trends in cognitive sciences, 9*(6), 290-295. doi: http://dx.doi.org/10.1016/j.tics.2005.04.012

Haggard, P., & Clark, S. (2003). Intentional action: Conscious experience and neural prediction. *Consciousness and Cognition, 12*(4), 695-707. doi: http://dx.doi.org/10.1016/S1053-8100(03)00052-7

Haggard, P., Clark, S., & Kalogeras, J. (2002). Voluntary action and conscious awareness. [10.1038/nn827]. *Nat Neurosci, 5*(4), 382-385.

Haggard, P., Martin, F., Taylor-Clarke, M., Jeannerod, M., & Franck, N. (2003). Awareness of action in schizophrenia. *NeuroReport, 14*(7), 1081-1085 1010.1097/1001.wnr.0000073684.0000000308.c0000073680.

Howes, O. D., & Kapur, S. (2009). The Dopamine Hypothesis of Schizophrenia: Version III—The Final Common Pathway. *Schizophrenia Bulletin, 35*(3), 549-562. doi: 10.1093/schbul/sbp006

Hughes, G., & Waszak, F. (2011). ERP correlates of action effect prediction and visual sensory attenuation in voluntary action. *NeuroImage, 56*(3), 1632-1640. doi: http://dx.doi.org/10.1016/j.neuroimage.2011.02.057

Hume, D. (1758). *Essays and treatises on several subjects*: Millar.

Humphreys, G. R., & Buehner, M. J. (2009). Magnitude estimation reveals temporal binding at super-second intervals. *Journal of Experimental Psychology: Human Perception and Performance, 35*(5), 1542-1549. doi: 10.1037/a0014492

Iglesias, S., Mathys, C., Brodersen, Kay H., Kasper, L., Piccirelli, M., den Ouden, Hanneke E. M., & Stephan, Klaas E. (2013). Hierarchical Prediction Errors in Midbrain and Basal Forebrain during Sensory Learning. *Neuron, 80*(2), 519-530. doi: http://dx.doi.org/10.1016/j.neuron.2013.09.009

Jahanshahi, M., Jones, C. R. G., Dirnberger, G., & Frith, C. D. (2006). The Substantia Nigra Pars Compacta and Temporal Processing. *The Journal of Neuroscience, 26*(47), 12266-12273. doi: 10.1523/jneurosci.2540-06.2006

Karson, C. N. (1983). Spontaneous eye-blink rates and dopaminergic systems. *Brain, 106*(3), 643-653. doi: 10.1093/brain/106.3.643

Kilner, J. M., Friston, K. J., & Frith, C. D. (2007). Predictive coding: an account of the mirror neuron system. *Cognitive Processing, 8*(3), 159-166. doi: 10.1007/s10339-007-0170-2

Kleiner, M., Brainard, D., Pelli, D., Ingling, A., Murray, R., & Broussard, C. (2007). What's new in Psychtoolbox-3. *Perception, 36*(14), 1.1-16.

Kühn, S., Brass, M., & Haggard, P. (2013). Feeling in control: Neural correlates of experience of agency. *Cortex, 49*(7), 1935-1942. doi: http://dx.doi.org/10.1016/j.cortex.2012.09.002

Körding, K. P., Beierholm, U., Ma, W. J., Quartz, S., Tenenbaum, J. B., & Shams, L. (2007). Causal Inference in Multisensory Perception. *PLoS ONE, 2*(9), e943. doi: 10.1371/journal.pone.0000943

Lancaster, J. L., Woldorff, M. G., Parsons, L. M., Liotti, M., Freitas, C. S., Rainey, L., . . . Fox, P. T. (2000). Automated Talairach Atlas labels for functional brain mapping.

Human Brain Mapping, 10(3), 120-131. doi: 10.1002/1097-0193(200007)10:3<120::aid-hbm30>3.0.co;2-8

Libet, B., Gleason, C. A., Wrigth, E. W., & Pearl, D. K. (1983). Time of conscious intention to act in relation to onset of cerebral activity (readiness-potential): The unconscious initiation of a freely voluntary act. *Brain, 106*(3), 623-642. doi: 10.1093/brain/106.3.623

Lieberman, M. D., & Cunningham, W. A. (2009). Type I and Type II error concerns in fMRI research: re-balancing the scale. *Social Cognitive and Affective Neuroscience, 4*(4), 423-428. doi: 10.1093/scan/nsp052

Lindner, A., Thier, P., Kircher, T. T. J., Haarmeier, T., & Leube, D. T. (2005). Disorders of Agency in Schizophrenia Correlate with an Inability to Compensate for the Sensory Consequences of Actions. *Current Biology, 15*(12), 1119-1124. doi: http://dx.doi.org/10.1016/j.cub.2005.05.049

Mackintosh, N. J. (1975). A theory of attention: Variations in the associability of stimuli with reinforcement 4. 82

Maldjian, J. A., Laurienti, P. J., Kraft, R. A., & Burdette, J. H. (2003). An automated method for neuroanatomic and cytoarchitectonic atlas-based interrogation of fMRI data sets. *NeuroImage, 19*(3), 1233-1239. doi: http://dx.doi.org/10.1016/S1053-8119(03)00169-1

Mikl, M., Mareček, R., Hluštík, P., Pavlicová, M., Drastich, A., Chlebus, P., . . . Krupa, P. (2008). Effects of spatial smoothing on fMRI group inferences. *Magnetic Resonance Imaging, 26*(4), 490-503. doi: http://dx.doi.org/10.1016/j.mri.2007.08.006

Miller, R. R., Barnet, R. C., & Grahame, N. J. (1995). Assessment of the Rescorla-Wagner model. *Psychological Bulletin, 117*(3), 363-386. doi: 10.1037/0033-2909.117.3.363

Moore, J. W., Dickinson, A., & Fletcher, P. C. (2011). Sense of agency, associative learning, and schizotypy. *Consciousness and Cognition, 20*(3), 792-800. doi: http://dx.doi.org/10.1016/j.concog.2011.01.002

Moore, J. W., & Fletcher, P. C. (2012). Sense of agency in health and disease: A review of cue integration approaches. *Consciousness and Cognition, 21*(1), 59-68. doi: http://dx.doi.org/10.1016/j.concog.2011.08.010

Moore, J. W., & Haggard, P. (2008). Awareness of action: Inference and prediction. *Consciousness and Cognition, 17*(1), 136-144. doi: http://dx.doi.org/10.1016/j.concog.2006.12.004

Moore, J. W., & Obhi, S. S. (2012). Intentional binding and the sense of agency: A review. *Consciousness and Cognition, 21*(1), 546-561. doi: http://dx.doi.org/10.1016/j.concog.2011.12.002

Moore, J. W., Ruge, D., Wenke, D., Rothwell, J., & Haggard, P. (2010). Disrupting the experience of control in the human brain: pre-supplementary motor area contributes to the sense of agency. *Proceedings of the Royal Society B: Biological Sciences, 277*(1693), 2503-2509. doi: 10.1098/rspb.2010.0404

Moore, J. W., Schneider, S. A., Schwingenschuh, P., Moretto, G., Bhatia, K. P., & Haggard, P. (2010). Dopaminergic medication boosts action–effect binding in Parkinson's disease. *Neuropsychologia, 48*(4), 1125-1132. doi: http://dx.doi.org/10.1016/j.neuropsychologia.2009.12.014

Moore, J. W., Wegner, D. M., & Haggard, P. (2009). Modulating the sense of agency with external cues. *Consciousness and Cognition, 18*(4), 1056-1064. doi: http://dx.doi.org/10.1016/j.concog.2009.05.004

Murray, G. K., Corlett, P. R., Clark, L., Pessiglione, M., Blackwell, A. D., Honey, G., . . . Fletcher, P. C. (2007). Substantia nigra/ventral tegmental reward prediction error disruption in psychosis. *Mol Psychiatry, 13*(3), 267-276. doi: http://www.nature.com/mp/journal/v13/n3/suppinfo/4002058s1.html

Pearce, J. M., & Hall, G. (1980). A model for Pavlovian learning: Variations in the effectiveness of conditioned but not of unconditioned stimuli. *Psychological Review, 87*(6), 532-552. doi: 10.1037/0033-295x.87.6.532

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spatial Vision, 10*(4), 437-442. doi: doi:10.1163/156856897X00366

Pessiglione, M., Seymour, B., Flandin, G., Dolan, R. J., & Frith, C. D. (2006). Dopamine-dependent prediction errors underpin reward-seeking behaviour in humans. [10.1038/nature05051]. *Nature, 442*(7106), 1042-1045. doi: http://www.nature.com/nature/journal/v442/n7106/suppinfo/nature05051_S1.html

Poldrack, R. A. (2007). Region of interest analysis for fMRI. *Social Cognitive and Affective Neuroscience, 2*(1), 67-70. doi: 10.1093/scan/nsm006

Poldrack, R. A., Fletcher, P. C., Henson, R. N., Worsley, K. J., Brett, M., & Nichols, T. E. (2008). Guidelines for reporting an fMRI study. *NeuroImage, 40*(2), 409-414. doi: http://dx.doi.org/10.1016/j.neuroimage.2007.11.048

Poonian, S. K., & Cunnington, R. (2013). Intentional binding in self-made and observed actions. *Experimental Brain Research,* 1-9. doi: 10.1007/s00221-013-3505-5

Raichle, M. E., MacLeod, A. M., Snyder, A. Z., Powers, W. J., Gusnard, D. A., & Shulman, G. L. (2001). A default mode of brain function. *Proceedings of the National Academy of Sciences, 98*(2), 676-682. doi: 10.1073/pnas.98.2.676

Rescorla, R. A., & Wagner, A. D. (1972). *A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement*. New York: Appleton-Century-Crofts.

Roussel, C., Hughes, G., & Waszak, F. (2013). A preactivation account of sensory attenuation. *Neuropsychologia, 51*(5), 922-929. doi: http://dx.doi.org/10.1016/j.neuropsychologia.2013.02.005

Salgado-Pineda, P., Fakra, E., Delaveau, P., McKenna, P. J., Pomarol-Clotet, E., & Blin, O. (2011). Correlated structural and functional brain abnormalities in the default mode network in schizophrenia patients. *Schizophrenia Research, 125*(2–3), 101-109. doi: http://dx.doi.org/10.1016/j.schres.2010.10.027

Sato, A. (2009). Both motor prediction and conceptual congruency between preview and action-effect contribute to explicit judgment of agency. *Cognition, 110*(1), 74-83. doi: http://dx.doi.org/10.1016/j.cognition.2008.10.011

Sato, A., & Yasuda, A. (2005). Illusion of sense of self-agency: discrepancy between the predicted and actual sensory consequences of actions modulates the sense of self-agency, but not the sense of self-ownership. *Cognition, 94*(3), 241-255. doi: http://dx.doi.org/10.1016/j.cognition.2004.04.003

Scholl, B. J., & Nakayama, K. (2002). Causal Capture: Contextual Effects on the Perception of Collision Events. *Psychological Science, 13*(6), 493-498. doi: 10.1111/1467-9280.00487

Scholl, B. J., & Tremoulet, P. D. (2000). Perceptual causality and animacy. *Trends in cognitive sciences, 4*(8), 299-309. doi: http://dx.doi.org/10.1016/S1364-6613(00)01506-0

Schultz, W., Dayan, P., & Montague, P. R. (1997). A Neural Substrate of Prediction and Reward. *Science, 275*(5306), 1593-1599. doi: 10.1126/science.275.5306.1593

Shergill, S. S., White, T. P., Joyce, D. W., Bays, P. M., Wolpert, D. M., & Frith, C. D. (2013). Modulation of somatosensory processing by action. *NeuroImage, 70*(0), 356-362. doi: http://dx.doi.org/10.1016/j.neuroimage.2012.12.043

Sladky, R., Friston, K. J., Tröstl, J., Cunnington, R., Moser, E., & Windischberger, C. (2011). Slice-timing effects and their correction in functional MRI. *NeuroImage, 58*(2), 588-594. doi: http://dx.doi.org/10.1016/j.neuroimage.2011.06.078

Spanagel, R., & Weiss, F. (1999). The dopamine hypothesis of reward: past and current status. *Trends in Neurosciences, 22*(11), 521-527. doi: http://dx.doi.org/10.1016/S0166-2236(99)01447-2

Storey, J. D. (2002). A direct approach to false discovery rates. *Journal of the Royal Statistical Society: Series B (Statistical Methodology), 64*(3), 479-498. doi: 10.1111/1467-9868.00346

Synofzik, M., Thier, P., Leube, D. T., Schlotterbeck, P., & Lindner, A. (2010). Misattributions of agency in schizophrenia are based on imprecise predictions about the sensory consequences of one's actions. *Brain, 133*(1), 262-271. doi: 10.1093/brain/awp291

Toda, M., & Abi-Dargham, A. (2007). Dopamine hypothesis of schizophrenia: Making sense of it all. *Current Psychiatry Reports, 9*(4), 329-336. doi: 10.1007/s11920-007-0041-7

Triantafyllou, C., Hoge, R. D., & Wald, L. L. (2006). Effect of spatial smoothing on physiological noise in high-resolution fMRI. *NeuroImage, 32*(2), 551-557. doi: http://dx.doi.org/10.1016/j.neuroimage.2006.04.182

Tsakiris, M., & Haggard, P. (2003). Awareness of somatic events associated with a voluntary action. *Experimental Brain Research, 149*(4), 439-446. doi: 10.1007/s00221-003-1386-8

Turk-Browne, N. B., Jungé, J. A., & Scholl, B. J. (2005). The Automaticity of Visual Statistical Learning. *Journal of Experimental Psychology: General, 134*(4), 552-564. doi: 10.1037/0096-3445.134.4.552

Turk-Browne, N. B., Scholl, B. J., Chun, M. M., & Johnson, M. K. (2008). Neural Evidence of Statistical Learning: Efficient Detection of Visual Regularities Without Awareness. *Journal of Cognitive Neuroscience, 21*(10), 1934-1945. doi: 10.1162/jocn.2009.21131

Turk-Browne, N. B., Scholl, B. J., Johnson, M. K., & Chun, M. M. (2010). Implicit Perceptual Anticipation Triggered by Statistical Learning. *The Journal of Neuroscience, 30*(33), 11177-11187. doi: 10.1523/jneurosci.0858-10.2010

Utevsky, A. V., Smith, D. V., & Huettel, S. A. (2014). Precuneus Is a Functional Core of the Default-Mode Network. *The Journal of Neuroscience, 34*(3), 932-940. doi: 10.1523/jneurosci.4227-13.2014

van der Post, J., de Waal, P. P., de Kam, M. L., Cohen, A. F., & van Gerven, J. M. A. (2004). No evidence of the usefulness of eye blinking as a marker for central dopaminergic activity. *Journal of Psychopharmacology, 18*(1), 109-114. doi: 10.1177/0269881104042832

Voss, M., Ingram, J. N., Haggard, P., & Wolpert, D. M. (2006). Sensorimotor attenuation by central motor command signals in the absence of movement. [10.1038/nn1592]. *Nat Neurosci, 9*(1), 26-27. doi: http://www.nature.com/neuro/journal/v9/n1/suppinfo/nn1592_S1.html

Voss, M., Moore, J., Hauser, M., Gallinat, J., Heinz, A., & Haggard, P. (2010). Altered awareness of action in schizophrenia: a specific deficit in predicting action consequences. *Brain, 133*(10), 3104-3112. doi: 10.1093/brain/awq152

Wagenmakers, E.-J. (2007). A practical solution to the pervasive problems ofp values. *Psychonomic Bulletin & Review, 14*(5), 779-804. doi: 10.3758/bf03194105

Wegner, D. M. (2003). The mind's best trick: how we experience conscious will. *Trends in cognitive sciences, 7*(2), 65-69. doi: http://dx.doi.org/10.1016/S1364-6613(03)00002-0

Williams, L. E., Ramachandran, V. S., Hubbard, E. M., Braff, D. L., & Light, G. A. (2010). Superior size–weight illusion performance in patients with schizophrenia: Evidence for deficits in forward models. *Schizophrenia Research, 121*(1–3), 101-106. doi: http://dx.doi.org/10.1016/j.schres.2009.10.021

Wise, R. A., & Rompre, P. P. (1989). Brain Dopamine and Reward. *Annual Review of Psychology, 40*(1), 191-225. doi: doi:10.1146/annurev.ps.40.020189.001203

Wohlschläger, A., Haggard, P., Gesierich, B., & Prinz, W. (2003). The Perceived Onset Time of Self- and Other-Generated Actions. *Psychological Science, 14*(6), 586-591. doi: 10.1046/j.0956-7976.2003.psci_1469.x

Wolpert, D. M. (1997). Computational approaches to motor control. *Trends in cognitive sciences, 1*(6), 209-216. doi: http://dx.doi.org/10.1016/S1364-6613(97)01070-X

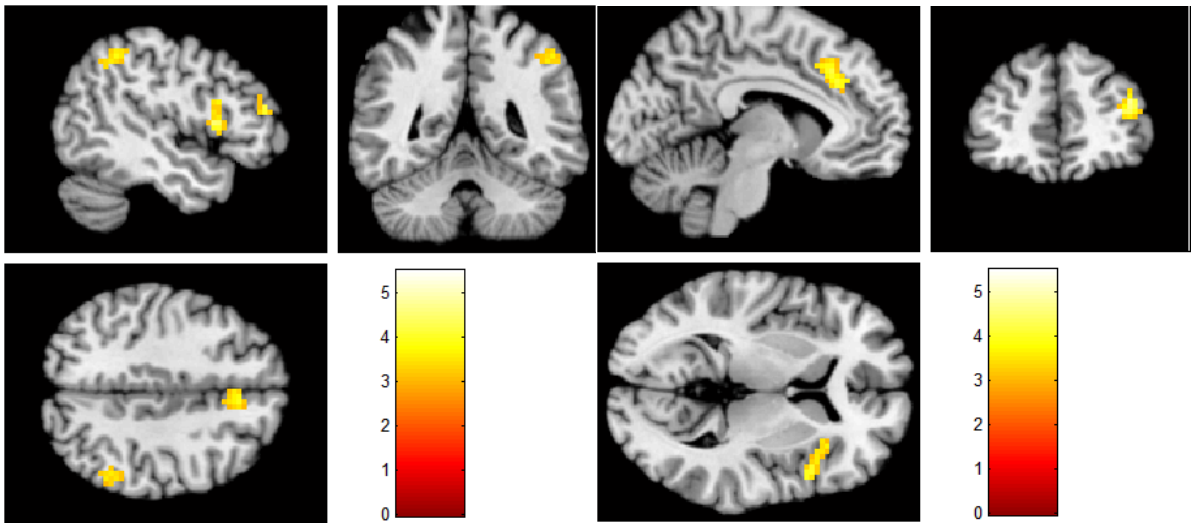
Wolpert, D. M., & Ghahramani, Z. (2000). Computational principles of movement neuroscience. *Nat Neurosci*.

6 Appendix

Supplementary table 1. Definition of variables in the individual regression analyses on the behavioural data.

Regressor name	**Variable type**	**Explanation**	**Definition**
time	continuous	Represents a linear increase in time according to number of trials performed.	Trial-wise increasing numbers between 0 and 1.
condition	categorical, dichotomous	Represents cue associated with colour outcome versus grey.	Button press associated with colour is defined as 1. The other button press is 0.
outcome	categorical, dichotomous	Actual outcome of the current button press in terms of colour or grey.	Colour is defined as 1 and grey as 0.
prediction	continuous	Stimulus-prediction regarding outcome.	Defined with respect to colour for button press with high probabilistic association to colour and with respect to grey for the other button press. Trial-specific values are estimated using the RW model.
surprise	categorical, dichotomous	Surprising outcome with respect to current prediction.	Surprising outcome is defined as 1 and expected outcome as 0.
prediction × *surprise*	continuous	Interaction between current prediction level and the violation (surprise) or fulfilment (expectancy) of that prediction.	Interaction term created by the element-wise multiplication of *prediction* and *surprise* regressors.

Supplementary figure 1. Indications of activity associated with increased rTB (uncorrected values).



The figures are showing t-map (p_{uncorr} = 0.005, k > 10) with clusters that did not survive correction for multiple comparisons for the contrast showing positive correlation with rTB. Peak activations are located in middle frontal gyrus (MNI 42, 41, 13; $p_{uncorr} = 5.11 \times 10^{-5}$; Z = 3.89), inferior frontal gyrus (MNI 48, 11, 7; $p_{uncorr} = 3.07 \times 10^{-4}$; Z = 3.43) extending into insula (MNI 33, 23, 1; $p_{uncorr} = 1.26 \times 10^{-3}$; Z = 3.02), middle cingulum (MNI 6, 26, 40; $p_{uncorr} = 3.39 \times 10^{-4}$; Z = 3.40), and inferior parietal lobe, (MNI 51, -46, 49; $p_{uncorr} = 1.06 \times 10^{-3}$; Z = 3.07). Activation maps are overlaid on the colin27 template brain.